matics Library

NOV 2 8 1945

AMERICAN JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

EDITED BY

R. BRAUER
UNIVERSITY OF TORONTO

H. WEYL
THE INSTITUTE FOR ADVANCED STUDY

F. D. MURNAGHAN
THE JOHNS HOPKINS UNIVERSITY

H. WHITNEY
HARVARD UNIVERSITY

A. WINTNER
THE JOHNS HOPKINS UNIVERSITY

WITH THE COÖPERATION OF

V. G. GROVE
M. H. HEINS
D. C. LEWIS
T. RADÓ
H. S. WALL

R. BAER
J. DOUGLAS
W. HUREWICZ
N. LEVINSON
G. PALL

S. BOCHNER
D. MONTGOMERY
C. L. SIEGEL
D. C. SPENCER
M. A. ZORN

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

Volume LXVII, Number 4
OCTOBER, 1945

THE JOHNS HOPKINS PRESS
BALTIMORE 18, MARYLAND
U. S. A.

CONTENTS

PAGE

On the representation of a group of order g in the field of the g-th roots of unity. By RICHARD BRAUER, ... 461

Ein Satz ueber quadratische Formen mit komplexen Koeffizienten. By ISSAI SCHUR, ... 472

Mean-values of arithmetical representations. By AUREL WINTNER, . 481

Dirichlet series meromorphic in a half-plane. By H. D. BRUNK, . 486

Extensions of plane continua mappings. By G. T. WHYBURN, . . 505

A theory of critical points. By MAGNUS R. HESTENES, ... 521

The four-vertex theorem for surfaces of constant curvature. By S. B. JACKSON, ... 563

Conformal classes of surfaces. By AARON FIALKOW, ... 583

The AMERICAN JOURNAL OF MATHEMATICS will appear four times yearly.

The subscription price of the JOURNAL for the current volume is $7.50 (foreign postage 50 cents); single numbers $2.00.

A few complete sets of the JOURNAL remain on sale.

Papers intended for publication in the JOURNAL may be sent to any of the Editors.

Editorial communications may be sent to Professor F. D. MURNAGHAN at The Johns Hopkins University.

Subscriptions to the JOURNAL and all business communications should be sent to THE JOHNS HOPKINS PRESS, BALTIMORE 18, MARYLAND, U. S. A.

Entered as second-class matter at the Baltimore, Maryland, Postoffice, acceptance for mailing at special rate of postage provided for in Section 1103, Act of October 3, 1917, Authorized on July 3, 1918.

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

ON THE REPRESENTATION OF A GROUP OF ORDER g IN THE FIELD OF THE g-TH ROOTS OF UNITY.*

By RICHARD BRAUER.

To Hermann Weyl on his sixtieth birthday.

Introduction. It has long been surmised that every irreducible representation $\mathfrak{L}$ of a group $\mathfrak{G}$ of order g can be written in the field of the g-th roots of unity. This question has attracted the efforts of a number of mathematicians. Improving an earlier result of H. Maschke [**20**],[1] W. Burnside [**14**] proved that this conjecture is actually true for a very extensive class of representations. I. Schur [**22**] showed that it is correct for all soluble groups. The case of representations of odd degree with real character was treated by A. Speiser [**25**].[2] The most important progress after that was made by H. Hasse [**11**] who showed that the representation $\mathfrak{L}$ can be written in the field of the g^n-th roots of unity where n is a suitable positive rational integer depending on $\mathfrak{L}$.

In the present paper, the old question is solved in its full generality for the first time. We prove the theorem:

THEOREM. *If $\mathfrak{G}$ is a group of finite order g, then every irreducible representation $\mathfrak{L}$ of $\mathfrak{G}$ can be written in the field Ω of the g-th roots of unity (that is, $\mathfrak{L}$ is similar to a representation with coefficients in Ω).*

The proof is obtained by combining the methods of I. Schur and of H. Hasse with results concerning the modular representations of groups.[3]

1. Remarks on notation. In the following, $\mathfrak{G}$ will denote a group of finite order g. If G is a fixed element of $\mathfrak{G}$, the elements of $\mathfrak{G}$ commuting with G form a subgroup $\mathfrak{N}(G)$, the *normalizer* of G. The class of conjugate

---

* Received August 19, 1945.

[1] Numbers in square brackets refer to the bibliography at the end.

[2] See also Schur [**24**], Brauer [**3**].

[3] For the definitions and theorems of this theory as far as they are used here, see Brauer-Nesbitt [**12**] and [**13**] Parts I and II, Nakayama, [**21**], Brauer [**7**], [**8**], [**9**], sections 1 and 4, [**10**], sections 1-4. Part I of [**13**] contains a short summary of the results and proofs of [**12**], with exception of the proof of Theorem I of [**12**]. Proofs of this theorem can also be found in [**7**], [**21**], and in Artin-Nesbitt-Thrall, [**2**], Chapter IX, 8.

elements containing the element G then consists of $g/n(G)$ elements where $n(G)$ is the order of $\mathfrak{N}(G)$.

A representation $\mathfrak{L}$ of $\mathfrak{G}$ is a group of matrices $\mathfrak{L}$ on which $\mathfrak{G}$ is mapped homomorphically; the coefficients of the matrices lie in a given field Λ. The image in $\mathfrak{L}$ of an element G of $\mathfrak{G}$ will be denoted by $\mathfrak{L}(G)$ and an analogous notation will be used throughout the paper. The word *representation* without further remark will always refer to a representation in a field of characteristic 0. In the other case, the term *modular representation* will be used. If p is the characteristic of the field in this case, then the elements R of $\mathfrak{G}$ whose order is prime to p are of importance; we denote them as the *p-regular elements* of $\mathfrak{G}$.

The group $\mathfrak{G}$ has k distinct absolutely irreducible[4] representations $\mathfrak{Z}_1, \mathfrak{Z}_2, \cdots, \mathfrak{Z}_k$ where k is the number of classes of conjugate elements in $\mathfrak{G}$. These representations can be chosen with coefficients in a suitable algebraic number field Λ of finite degree over the field P of rational numbers. If p is a given rational prime, and $\mathfrak{q}$ a prime ideal divisor of p in Λ, we may assume without restriction that all coefficients of every $\mathfrak{Z}_\kappa(G)$, G in $\mathfrak{G}$, are $\mathfrak{q}$-integers, that is, that they are quotients u/v of integers of Λ such that v is not divisible by $\mathfrak{q}$. If every coefficient of $\mathfrak{Z}_\kappa$ is replaced by its residue class modulo $\mathfrak{q}$, a modular representation $\mathfrak{Z}^*_\kappa$ in a field of characteristic p is obtained. The distinct constituents of $\mathfrak{Z}^*_1, \mathfrak{Z}^*_2, \cdots, \mathfrak{Z}^*_k$ furnish a complete system of absolutely irreducible modular representations $\mathfrak{F}_1, \mathfrak{F}_2, \cdots, \mathfrak{F}_l$ for the characteristic p. The numbers $d_{\kappa\lambda}$ indicating how often $\mathfrak{F}_\lambda$ appears as a constituent in $\mathfrak{Z}^*_\kappa$ are the *decomposition numbers* (for p); they are non-negative rational integers.

The character of $\mathfrak{Z}_\kappa$ will be denoted by ζ_κ and the character of $\mathfrak{F}_\lambda$ by ϕ_λ.[5] We then have

$$\zeta_\kappa(R) = \sum_{\lambda=1}^{l} d_{\kappa\lambda}\phi_\lambda(R) \tag{1}$$

for p-regular elements R of $\mathfrak{G}$.

2. Proof of the theorem for modular representations $\mathfrak{L}$. If $\mathfrak{L}$ is an irreducible modular representation $\mathfrak{F}$ of $\mathfrak{G}$ in a field Π of characteristic p, then Π will certainly contain the traces $\mathrm{tr}(\mathfrak{F}(G))$ of all the matrices $\mathfrak{F}(G)$. Let Π_0 be the field generated by these g traces. Since every trace is a sum of

[4] A representation is absolutely irreducible, if it remains irreducible in every extension field.

[5] It should be observed that the modular characters are defined as complex numbers, not as elements of a modular field, see for instance [13], § 6. The trace of the matrix $\mathfrak{F}_\lambda(R)$ for a p-regular element R of $\mathfrak{G}$ then is the residue class $\phi_\lambda(R)^*$ of $\phi_\lambda(R)$ (mod $\mathfrak{q}$).

g-th roots of unity, the field Π_0 is a finite Galois field. It follows [6] that the representation $\mathfrak{F}$ can be written in the field Π_0 itself. As Π_0 is a subfield of the field of the g-th roots of unity, this proves the theorem for modular representations $\mathfrak{L}$.

3. A connection between the $\mathfrak{p}$-indices [7] and the decomposition numbers. We now come to the much more difficult case of non-modular representations $\mathfrak{L}$. It will be necessary to derive first a number of lemmas.

Let K_0 be the field obtained by adjoining all the modular characters $\phi_1(R), \phi_2(R), \cdots, \phi_l(R)$ (for all p-regular elements R of $\mathfrak{G}$) to the field of rational numbers P,

$$K_0 = P(\phi_1, \phi_2, \cdots, \phi_l). \tag{2}$$

To K_0, adjoin the values $\zeta_\kappa(G)$ of a fixed character ζ_κ for all elements G of $\mathfrak{G}$,

$$K = K_0(\zeta_\kappa). \tag{3}$$

Let p be a rational prime number and let $\mathfrak{p}$ be a fixed prime ideal of K dividing p. If θ is any $\mathfrak{p}$-integer of K we denote its residue class (mod $\mathfrak{p}$) by θ^*. The totality of all θ^* forms the residue class field K* of K (mod $\mathfrak{p}$).

If R is a p-regular element of $\mathfrak{G}$, the trace of $\mathfrak{F}_\lambda(R)$ is equal to $\phi_\lambda(R)^*$.[8] If G is an arbitrary element of $\mathfrak{G}$, it may be written in the form $G = PR$ where the order of P is a power of p while R is a p-regular element belonging to $\mathfrak{N}(P)$. Then [9]

$$\mathrm{tr}(\mathfrak{F}_\lambda(G)) = \mathrm{tr}(\mathfrak{F}_\lambda(R)) = \phi_\lambda(R)^*.$$

In any case, the field K* contains the traces of all matrices $\mathfrak{F}_\lambda(\mathfrak{G})$. The result of Section 2 shows that $\mathfrak{F}_1, \mathfrak{F}_2, \cdots, \mathfrak{F}_l$ may be written with coefficients in K*.

It has been shown previously [10] that there exists a representation $\mathfrak{W}_\lambda$ of $\mathfrak{G}$

---

[6] Brauer [5] §4, p. 101. The fact used here is equivalent to the theorem of Wedderburn [27] that every division algebra over a finite field is commutative.

[7] The theory of the index of a representation with regard to a field was developed by I. Schur, [22], [23]. The connection of this theory with the theory of algebras was given in Brauer [4], [5]. The $\mathfrak{p}$-indices were introduced in Hasse [16], [17], Brauer-Hasse-Noether [11]. See also the presentation of the theory of algebras in the books Deuring [15], Albert [1], Weyl [28], Jacobson [18], van der Waerden [26], Artin-Nesbitt-Thrall [2], and of the arithmetic parts of the theory in the first two of these books.

[8] See footnote 5.

[9] See for instance [13], § 6.

[10] Brauer [7], § 6. See also the arrangement of the proofs in Artin-Nesbitt-Thrall [2], Chapter IX, 8. An equivalent theorem is given in Nakayama [21], § 4.

with coefficients in the $\mathfrak{p}$-adic extension field $\mathbf{K}(\mathfrak{p})$ of $\mathbf{K}$ with the following properties

(a) $\mathfrak{W}_\lambda(G)$ has $\mathfrak{p}$-integral coefficients for every G in $\mathfrak{G}$.

(b) If every coefficient of $\mathfrak{W}_\lambda(G)$ is replaced by its residue class (mod $\mathfrak{p}$), a modular representation $\mathfrak{W}^*_\lambda$ is obtained which is an indecomposable constituent $\mathfrak{U}_\lambda$ of the modular regular representation and which has the irreducible bottom constituent $\mathfrak{F}_\lambda$. Here, λ can have any of the values $1, 2, \cdots, l$.

Since $\mathfrak{F}_\lambda$ is absolutely irreducible, the representation $\mathfrak{U}_\lambda$ will remain indecomposable in any extension field Λ^* of $\mathbf{K}^*$.[11] If $\mathbf{K}$ is replaced by an extension field Λ of finite degree, if $\mathfrak{p}$ is replaced by a prime ideal divisor $\mathfrak{q}$ of $\mathfrak{p}$, and if $\mathbf{K}(\mathfrak{p})$ is replaced by the $\mathfrak{q}$-adic extension field $\Lambda(\mathfrak{q})$, the representation $\mathfrak{W}_\lambda$ will retain the properties (a) and (b). For instance, as in Section **1**, we may take Λ as a field in which $\mathfrak{Z}_1, \mathfrak{Z}_2, \cdots, \mathfrak{Z}_k$ can be written. Then $\mathfrak{Z}_\kappa$ appears with the multiplicity $d_{\kappa\lambda}$ in $\mathfrak{W}_\lambda$,[12] and therefore,[13] the Schur index $\mu(\mathfrak{p})$ of $\mathfrak{Z}_\kappa$ with regard to $\mathbf{K}(\mathfrak{p})$ is a divisor of $d_{\kappa\lambda}$. This index is termed the $\mathfrak{p}$-*index* of $\mathfrak{Z}_\kappa$ with regard to $\mathbf{K}$.

We thus have

LEMMA 1. *If* $\mathbf{K}$ *is obtained from the field of rational numbers by adjoining the modular characters* $\phi_1, \phi_2, \cdots, \phi_l$ *and the character* ζ_κ *of the representation* $\mathfrak{Z}_\kappa$, *then the* $\mathfrak{p}$-*index* $\mu(\mathfrak{p})$ *of* $\mathfrak{Z}_\kappa$ *with regard to* $\mathbf{K}$ *divides all decomposition numbers* $d_{\kappa\lambda}$ *with the fixed first subscript* κ.

4. On the power of p dividing the $\mathfrak{p}$-index of $\mathfrak{Z}_\kappa$. Let p^a be the highest power of p dividing g, so that

$$g = p^a g', \qquad (g', p) = 1. \tag{4}$$

Every $\phi_\lambda(R)$ for p-regular R is a sum of g'-th roots of unity. Hence the field $\mathbf{K}_0$ in (2) is contained in the field $\mathbf{T}$ of the g'-th roots of unity. Now (3) implies

$$\mathbf{K} \subseteq \mathbf{T}(\zeta_\kappa). \tag{5}$$

---

[11] As $\mathfrak{F}_\lambda$ is absolutely irreducible, the degree of the corresponding indecomposable constituent of the regular representation is equal to the multiplicity of $\mathfrak{F}_\lambda$ in the second regular representation, see [7], § 2 or [2]. This shows that $\mathfrak{U}_\lambda$ can not be decomposable in Λ^*.

[12] Brauer [7], equation (14). The proof of Lemma 1 can also be obtained directly from [7], equation (11), which is identical with Artin-Nesbitt-Thrall [2], equation (8.6) in Chapter IX.

[13] Schur [22], Theorem IV.

On the other hand, if ϵ denotes a primitive p^a-th root of unity, then $\Omega = \mathbf{T}(\epsilon)$ is the field of the g-th roots of unity, and we have

$$\mathbf{T} \subseteq \mathbf{T}(\zeta_\kappa) \subseteq \Omega = \mathbf{T}(\epsilon). \tag{6}$$

The degree r of $\mathbf{T}(\zeta_\kappa)$ with regard to $\mathbf{T}$ is equal to the number of p-conjugate characters to ζ_κ (that is, the number of distinct characters into which ζ_κ can be carried by an automorphism of Ω leaving the g'-th roots of unity fixed). If z_κ is the degree of $\mathfrak{Z}_\kappa$ then, as shown previously,[14]

$$rz_\kappa \not\equiv 0 \pmod{p^{a+1}}, \tag{7}$$

$$rz_\kappa \not\equiv 0 \pmod{p^a}, \quad \text{if} \quad z_\kappa \not\equiv 0 \pmod{p^a}. \tag{8}$$

If $z_\kappa \equiv 0 \pmod{p^a}$, a suitable $d_{\kappa\lambda}$ has the value 1,[15] and Lemma 1 shows that $\mu(\mathfrak{p}) \not\equiv 0 \pmod{p}$.

Assume now that $z_\kappa \not\equiv 0 \pmod{p^a}$. Then (8) can be applied. The $\mathfrak{p}$-index $\mu(\mathfrak{p})$, which is the Schur index of a representation $\mathfrak{Z}_\kappa$, divides the degree z_κ of this representation.[16] Hence

$$r\mu(\mathfrak{p}) \not\equiv 0 \pmod{p^a}. \tag{9}$$

The field Ω of the g-th roots of unity has the degree

$$[\Omega : \mathbf{T}] = p^{a-1}(p-1)$$

over the field $\mathbf{T}$ of the g'-th roots of unity, cf. (4). As the degree of $\mathbf{T}(\zeta_\kappa)$ over $\mathbf{T}$ was r, it follows that the degree of Ω over $\mathbf{T}(\zeta_\kappa)$ is given by

$$[\Omega : \mathbf{T}(\zeta_\kappa)] = p^{a-1}(p-1)/r.$$

Let $\mathfrak{P}$ be a prime ideal divisor of $\mathfrak{p}$ in Ω and let $\mathfrak{r}$ and $\mathfrak{r}_1$ be the prime ideals of $\mathbf{T}$ and $\mathbf{T}(\zeta_\kappa)$, respectively, which are divisible by $\mathfrak{P}$. Then $p \equiv 0 \pmod{\mathfrak{r}}$, and it is well known that

$$\mathfrak{r} = \mathfrak{P}^h \quad \text{with} \quad h = [\Omega : \mathbf{T}].$$

The relation (6) implies

$$\mathfrak{r}_1 = \mathfrak{P}^j \quad \text{with} \quad j = [\Omega : \mathbf{T}(\zeta_\kappa)],$$

and it now follows from (5) that if $\mathfrak{P}^e$ is the highest power of $\mathfrak{P}$ dividing the

[14] Brauer [10], Theorem 2, p. 943.
[15] Brauer-Nesbitt [13], Theorem 1, p. 565.
[16] Schur [22], Theorem V.

prime ideal $\mathfrak{p}$, the exponent e is divisible by $j = [\Omega : \mathbf{T}(\zeta_\kappa)] = p^{a-1}(p-1)/r$.

The $\mathfrak{p}$-degree of Ω over $\mathbf{K}$ (that is the degree $n(\mathfrak{p})$ of the $\mathfrak{P}$-adic extension field $\Omega(\mathfrak{P})$ over $\mathbf{K}(\mathfrak{p})$),[17] is divisible by e. Hence $n(\mathfrak{p})$ is divisible by $p^{a-1}(p-1)/r$, and this gives

$$rn(\mathfrak{p}) \equiv 0 \pmod{p^{a-1}}. \tag{10}$$

On comparing (9) and (10), we obtain

LEMMA 2. *In the notation of Lemma* 1, *the highest power of* p *dividing* $\mu(\mathfrak{p})$ *also divides the* $\mathfrak{p}$*-degree* $n(\mathfrak{p})$ *of the field* Ω *of the* g*-th roots of unity over* $\mathbf{K}$.

5. A lemma concerning the index of $\mathfrak{Z}_\kappa$ with regard to Ω. Choose a system of elements

$$P_0 = 1, P_1, P_2, \cdots, P_s \tag{11}$$[18]

in $\mathfrak{G}$, such that (a) the order of every P_i is a power of p, (b) no two of these elements are conjugate in $\mathfrak{G}$, (c) every element of an order p^a is conjugate to one of the P_i. Set

$$\mathfrak{N}_i = \mathfrak{N}(P_i), \tag{12}$$

and determine a complete system of elements

$$V_1^i, V_2^i, \cdots \tag{13}$$

representing the different classes of p-regular elements in $\mathfrak{N}_i$. The elements $P_i V_j^i$ for all possible indices i, j form a system of representatives for the classes of conjugate elements in $\mathfrak{G}$.

The representation $\mathfrak{Z}_\kappa$ of $\mathfrak{G}$ induces a representation $\mathfrak{Z}_\kappa(\mathfrak{N}_i)$ of the subgroup $\mathfrak{N}_i$. If $\mathfrak{A}$ is an irreducible constituent, it represents the invariant element P_i of $\mathfrak{N}_i$ by a matrix of the form $\epsilon^\nu I$ where ϵ again denotes a primitive p^a-th root of unity. Hence

$$\mathfrak{A}(P_i V_j^i) = \mathfrak{A}(P_i)\mathfrak{A}(V_j^i) = \epsilon^\nu \mathfrak{A}(V_j^i)$$

which implies

$$\operatorname{tr} \mathfrak{A}(P_i V_j^i) = \epsilon^\nu \operatorname{tr} \mathfrak{A}(V_j^i). \tag{14}$$

The character $\operatorname{tr} \mathfrak{A}$ can be expressed by the irreducible modular characters $\phi_1^i, \phi_2^i, \cdots$ of $\mathfrak{N}_i$. We thus see that the left side in (14) is a linear com-

[17] See, for instance, Hasse [17], (3.3).

[18] For the following, cf. Brauer [9], Section 1.

bination of $\phi_1^i(V_j^i), \phi_2^i(V_j^i), \cdots$ with coefficients which are independent of j. These coefficients are non-negative rational integers multiplied by ϵ^ν.

Adding the formulae (14) for all irreducible constituents of $\mathfrak{Z}_\kappa(\mathfrak{N}_i)$, each taken with the correct multiplicity, we finally obtain formulae

$$\zeta_\kappa(P_i V_j^i) = \sum_\rho d_{\kappa\rho}^i \phi_\rho^i(V_j^i) \tag{15}$$

where the $d_{\kappa\rho}^i$ are algebraic integers, the *generalized decomposition numbers* of $\mathfrak{G}$.

We now prove

LEMMA 3. *Assume that all irreducible representations of $\mathfrak{N}_i = \mathfrak{N}(P_i)$ (for a fixed i) can be written in a field $\mathfrak{Z}$. The Schur index m of $\mathfrak{Z}_\kappa$ with regard to $\mathfrak{Z}$ divides $d_{\kappa\rho}^i$ for all values of ρ.*

Proof. If $\mathfrak{N}_i = \mathfrak{G}$, the assumption implies that $m = 1$, and the statement is trivial. If $\mathfrak{N}_i \subset \mathfrak{G}$, then it follows from a theorem of I. Schur,[19] that m divides the multiplicity of every constituent $\mathfrak{A}$ of $\mathfrak{Z}_\kappa(\mathfrak{N}_i)$. In deriving (15) from (14) we then have to multiply each equation (14) by a multiple of m before adding. This gives the statement of Lemma 3.

6. Proof of the theorem. Because of the complete reducibility of the representations of $\mathfrak{G}$, it is sufficient to prove the theorem formulated in the introduction for absolutely irreducible representations $\mathfrak{L}$. Assume that the theorem is false. Take a group $\mathfrak{G}$ of minimal order g for which there exists an absolutely irreducible representation $\mathfrak{L}$ which cannot be written in the field Ω of the g-th roots of unity. In other words, the Schur index m of $\mathfrak{L}$ with regard to Ω is different from 1. The representation $\mathfrak{L}$ is similar to one of the $\mathfrak{Z}_\kappa$; we may assume $\mathfrak{L} = \mathfrak{Z}_\kappa$ for a certain κ. This means [20] that there exists a prime ideal $\mathfrak{P}$ of Ω, finite or infinite, such that the $\mathfrak{P}$-index $\mathrm{M}(\mathfrak{P})$ of $\mathfrak{Z}_\kappa$ with regard to Ω is different from 1,

$$\mathrm{M}(\mathfrak{P}) > 1. \tag{16}$$

Since g must be larger than 2, the field Ω has no real conjugate. Then $\mathfrak{P}$ in (16) can not be infinite.

Assume that $\mathfrak{P}$ is finite. Let K have the same significance as in Lemma 1. Then $\mathrm{K} \subseteq \Omega$. Let $\mathfrak{p}$ be the prime ideal of K divisible by $\mathfrak{P}$, and again let p be the rational prime number divisible by $\mathfrak{p}$.

If for an element P_i in (11) we have $\mathfrak{N}_i \subset \mathfrak{G}$, the group $\mathfrak{N}_i$ has an order

[19] Schur [22], Theorem IX a. See also Brauer [6], Theorem IV.
[20] Brauer-Hasse-Noether [11], Theorem I.

which is a proper divisor of $\mathfrak{G}$. The theorem is correct for $\mathfrak{N}_i$, and all representations of $\mathfrak{N}_i$ can be written in Ω. Lemma 3 yields

$$d_{\kappa\lambda}{}^{i} \equiv 0 \pmod{m} \tag{17}$$

for all λ.

On the other hand, if $\mathfrak{N}(P_i) = \mathfrak{G}$ (as, for instance, for $i = 0$, $P_0 = 1$), the constituent $\mathfrak{A}$ of $\mathfrak{Z}_\kappa(\mathfrak{N}_i) = \mathfrak{Z}_\kappa(\mathfrak{G})$ in Section **5** coincides with $\mathfrak{Z}_\kappa$. Hence (14) reads

$$\zeta_\kappa(P_i V_j{}^{i}) = \epsilon^{\nu}\zeta_\kappa(V_j{}^{i}).$$

The equation (1) now yields

$$\zeta_\kappa(P_i V_j{}^{i}) = \epsilon^{\nu} \sum_{\rho} d_{\kappa\rho}\phi_\rho(V_j{}^{i}). \tag{18}$$

The irreducible modular characters $\phi_1{}^{i}, \phi_2{}^{i}, \cdots$ of $\mathfrak{N}_i = \mathfrak{G}$ are the characters $\phi_1, \phi_2, \cdots$; we may assume $\phi_\rho{}^{i} = \phi_\rho$. On comparing (15) and (18), we have $d_{\kappa\rho}{}^{i} = \epsilon^{\nu} d_{\kappa\rho}$. Now Lemma 1 shows that

$$d_{\kappa\rho}{}^{i} \equiv 0 \pmod{\mu(\mathfrak{p})}. \tag{19}$$

In any case, every $d_{\kappa\rho}{}^{i}$ with the fixed first subscript κ is divisible by the greatest common divisor of $\mu(\mathfrak{p})$ and m.

The orthogonality relation for group characters gives

$$(1/g) \sum \zeta_\kappa(G)\overline{\zeta_\kappa(G)} = 1 \tag{20}$$

where G ranges over all elements of $\mathfrak{G}$. The elements $P_i V_j{}^{i}$ form a complete system of representatives for the classes of conjugate elements in $\mathfrak{G}$. The class of $P_i V_j{}^{i}$ contains $g/n(P_i V_j{}^{i})$ elements. Now (20) can be written in the form

$$\sum_i \sum_j \zeta_\kappa(P_i V_j{}^{i})\overline{\zeta_\kappa(P_i V_j{}^{i})}/n(P_i V_j{}^{i}) = 1.$$

Combining this with (15), we have

$$\sum_i \sum_j \sum_\rho \sum_\sigma d_{\kappa\rho}{}^{i}\overline{d_{\kappa\sigma}{}^{i}}\phi_\rho{}^{i}(V_j{}^{i})\overline{\phi_\sigma{}^{i}(V_j{}^{i})}/n(P_i V_j{}^{i}) = 1. \tag{21}$$

In the group $\mathfrak{N}_i$ of order $n(P_i)$, the number of elements in the class of $V_j{}^{i}$ is $n(P_i)/n(P_i V_j{}^{i})$ since every element of $\mathfrak{G}$ commuting with $P_i V_j{}^{i}$ automatically lies in $\mathfrak{N}_i = \mathfrak{N}(P_i)$. Hence the orthogonality relation for the modular group characters [21] of $\mathfrak{N}_i$ reads

$$\sum_j \phi_\rho{}^{i}(V_j{}^{i})\overline{\phi_\sigma{}^{i}(V_j{}^{i})}/n(P_i V_j{}^{i}) = \gamma_{\rho\sigma}{}^{i} \tag{22}$$

---

[21] Brauer-Nesbitt **[13]**, (21).

where the matrix $(\gamma_{\rho\sigma}{}^{i})$ is the inverse of the matrix $C^{(i)}$ of the Cartan-invariants of $\mathfrak{N}_i$.

On account of (22), the equation (21) can be written in the form

$$\sum_{i}\sum_{\rho}\sum_{\sigma} d_{\kappa\rho}{}^{i}\overline{d_{\kappa\sigma}{}^{i}}\gamma_{\rho\sigma}{}^{i} = 1.$$

For a suitable rational integer t the products $p^{t}\gamma_{\rho\sigma}{}^{i}$ are rational integers.[22] We write the last equation in the form

$$\sum_{i}\sum_{\rho}\sum_{\sigma} d_{\kappa\rho}{}^{i}\overline{d_{\kappa\sigma}{}^{i}}(p^{t}\gamma_{\rho\sigma}{}^{i}) = p^{t}. \tag{23}$$

Since every $d_{\kappa\rho}{}^{i}$ was divisible by the greatest common divisor of $\mu(\mathfrak{p})$ and m, it follows, from (23), that the square of this greatest common divisor divides p^{t}. In any case, this greatest common divisor $(m, \mu(\mathfrak{p}))$ is a power of p.

The $\mathfrak{P}$-index $\mathrm{M}(\mathfrak{P})$ of $\mathfrak{Z}_\kappa$ with regard to Ω divides both the $\mathfrak{p}$-index of $\mathfrak{Z}_\kappa$ with regard to K and the index m of $\mathfrak{Z}_\kappa$ with regard to Ω. Hence

$$\mathrm{M}(\mathfrak{P}) = p^{b} \tag{24}$$

for a certain integral rational exponent $b \geqq 0$.

The absolutely irreducible representation $\mathfrak{Z}_\kappa$ has the index $\mu(\mathfrak{p})$ with respect to the field $\mathrm{K}(\mathfrak{p})$ which contains its character ζ_κ. Then $\mathfrak{Z}_\kappa$ determines a normal simple algebra A of index $\mu(\mathfrak{p})$ over $\mathrm{K}(\mathfrak{p})$ which consists of all linear combinations of elements of $\mathfrak{Z}_\kappa$ with coefficients in $\mathrm{K}(\mathfrak{p})$. The Hasse invariant [23] of A is a rational number $\rho = q/\mu(\mathfrak{p})$ with the denominator $\mu(\mathfrak{p})$, $(q, \mu(\mathfrak{p})) = 1$. If $\mathrm{K}(\mathfrak{p})$ is replaced by the extension field $\Omega(\mathfrak{P})$, then ρ is multiplied [24] by the relative degree $n(\mathfrak{p}) = [\Omega(\mathfrak{P}) : \mathrm{K}(\mathfrak{p})]$. Therefore, the new index $M(\mathfrak{P})$ is the denominator of $\rho\, n(\mathfrak{p}) = qn(\mathfrak{p})/\mu(\mathfrak{p})$ after numerator and denominator have been freed of their common divisor. Hence

$$\mathrm{M}(\mathfrak{P}) = \mu(\mathfrak{p})/(\mu(\mathfrak{p}), n(\mathfrak{p})). \tag{25}$$

Combination of (25) with Lemma 2 shows that $\mathrm{M}(\mathfrak{P})$ is not divisible by p, and (24) then implies that $\mathrm{M}(\mathfrak{P}) = 1$, in contradiction to (16). Hence, the assumption $m \neq 1$ leads to a contradiction, and the theorem is proved.

---

[22] The coefficients of $C^{(i)}$ are rational integers; the determinant of $C^{(i)}$ is a power of p, Brauer [8]. This shows already that $\gamma_{\rho\sigma}{}^{i}$ is a rational number whose denominator is a power of p. As can be seen easily, t can be chosen equal to a, but this is not needed in the following.

[23] Hasse [17], (5.1). See also Deuring [15], Chapter VII, §§ 2, 4.

[24] Köthe [19], Hasse [17], (5.2), also Deuring [15], Chapter VII, § 2.

7. The case of reducible representations. The first remark in Section 6 shows that the theorem also holds for reducible ordinary representations $\mathfrak{L}$. However, this is no longer true in the case of modular representations. For instance, if Π_0 denotes the Galois field with two elements and if t is a transcendental element with regard to Π_0 the four matrices

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ t & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ t+1 & 0 & 1 \end{pmatrix}$$

form a group of order four which can not be written in Π_0 as can be shown without difficulty.

UNIVERSITY OF TORONTO.

BIBLIOGRAPHY.

[1]. A. A. Albert, *Structure of Algebras*, American Mathematical Society Colloquium Publications, vol. 24 (1939).

[2]. E. Artin, C. J. Nesbitt, and R. M. Thrall, "Rings with minimum condition," *University of Michigan Publications in Mathematics*, No. 1, Ann Arbor, 1944.

[3]. R. Brauer, "Ueber Zusammenhänge zwischen arithmetischen und invariantentheoretischen Eigenschaften von Gruppen linearer Substitutionen," *Sitz. Ber. Preuss. Akad. Wiss.* (1926), pp. 410-416.

[4]. R. Brauer, "Untersuchungen über die arithmetischen Eigenschaften von Gruppen linearer Substitutionen, I," *Mathematische Zeitschrift*, vol. 28 (1928), pp. 677-696.

[5]. R. Brauer, "Ueber Systeme hyperkomplexer Zahlen," *Mathematische Zeitschrift*, vol. 30 (1929), pp. 79-107.

[6]. R. Brauer, "Untersuchungen über die arithmetischen Eigenschaften von Gruppen linearer Substitutionen, II," *Mathematische Zeitschrift*, vol. 31 (1930), pp. 737-747.

[7]. R. Brauer, "On modular and p-adic representations of algebras," *Proceedings of the National Academy of Sciences*, vol. 25 (1939), pp. 252-258.

[8]. R. Brauer, "On the Cartan invariants of groups of finite order," *Annals of Mathematics*, vol. 42 (1941), pp. 53-61.

[9]. R. Brauer, "On the connection between the ordinary and the modular characters of groups of finite order," *Annals of Mathematics*, vol. 42 (1941), pp. 926-935.

[10]. R. Brauer, "Investigations on group characters," *Annals of Mathematics*, vol. 42 (1941), pp. 936-958.

[11]. R. Brauer, H. Hasse, and E. Noether, "Beweis eines Hauptsatzes in der Theorie der Algebren," *Jour. reine angew. Math.*, vol. 167 (1932), pp. 399-404.

[12]. R. Brauer and C. J. Nesbitt, *On the Modular Representations of Groups of Finite Order*, Toronto Studies, 1937.

[13]. R. Brauer and C. J. Nesbitt, "On the modular characters of groups," *Annals of Mathematics*, vol. 42 (1941), pp. 556-590.

[14]. W. Burnside, "On the complete reduction of any transitive permutation group and on the arithmetic nature of the coefficients in its irreducible components," *Proceedings of the London Mathematical Society* (2), vol. 3 (1905), pp. 239-252.

[15]. M. Deuring, "Algebren," *Ergebnisse der Math.*, vol. 4 (Berlin, 1935).

[16]. H. Hasse, "Theory of cyclic algebras over an algebraic number field," *Transactions of the American Mathematical Society*, vol. 34 (1932), pp. 171-214, and pp. 727-730.

[17]. H. Hasse, "Die Struktur der R. Brauerschen Algebrenklassengruppe über einem algebraischen Zahlkörper," *Mathematische Annalen*, vol. 107 (1933), pp. 731-760.

[18]. N. Jacobson, "The Theory of Rings," *Mathematical Surveys*, vol. II, 1943.

[19]. G. Käthe, "Erweiterung des Zeutrums einfacher Algebren," *Mathematische Annalen*, vol. 107 (1933), pp. 761-766.

[20]. H. Maschke, "Ueber den arithmetischen Charakter der Coefficienten der Substitutionen endlicher linearer Substitutionsgruppen," *Mathematische Annalen*, vol. 50 (1898), pp. 482-498.

[21]. T. Nakayama, "Some studies on regular representations, induced representations, and modular representations," *Annals of Mathematics*, vol. 39 (1938), pp. 361-369.

[22]. I. Schur, "Arithmetische Untersuchungen über endliche Gruppen linearer Substitutionen," *Sitz. Ber. Preuss. Akad. Wiss.*, 1906, pp. 164-184.

[23]. I. Schur, "Beiträge zur Theorie der Gruppen linearer homogener Substitutionen," *Transactions of the American Mathematical Society*, vol. 15 (1909), pp. 159-175.

[24]. I. Schur, "Einige Bemerkungen zu der vorstehenden Arbeit des Herrn A. Speiser," *Mathematische Zeitschrift*, vol. 5 (1919), pp. 6-10.

[25]. A. Speiser, "Zahlentheoretische Sätze aus der Gruppentheorie," *Mathematische Zeitschrift*, vol. 5 (1919), pp. 1-6.

[26]. B. L. van der Waerden, *Moderne Algebra*, vol. II, 2nd edition, Berlin, 1940.

[27]. J. H. M. Wedderburn, "A theorem on finite algebras," *Transactions of the American Mathematical Society*, vol. 6 (1905), pp. 349-352.

[28]. H. Weyl, "The Classical Groups," Princeton, 1939.

EIN SATZ UEBER QUADRATISCHE FORMEN MIT KOMPLEXEN KOEFFIZIENTEN.*

By Issai Schur.

Im folgenden werde ich den Satz beweisen:

I. *Jede quadratische Form mit reellen oder komplexen Koeffizienten* $f(x) = \sum_{\kappa,\lambda=1}^{n} c_{\kappa\lambda} x_\kappa x_\lambda$ $(c_{\kappa\lambda} = c_{\lambda\kappa})$ *laesst sich in der Gestalt*

$$f(x) = \sum_{\nu=1}^{n} \omega_\nu \phi_\nu(x)^2, \qquad \phi_\nu(x) = \sum_{\alpha=1}^{n} d_{\nu\alpha} x_\alpha \tag{1}$$

darstellen, wobei $\omega_1, \cdots, \omega_n$ *nicht-negative reelle Zahlen sind und*

$$|\phi_1|^2 + |\phi_2|^2 + \cdots + |\phi_n|^2 = |x_1|^2 + |x_2|^2 + \cdots + |x_n|^2 \tag{2}$$

oder, was dasselbe ist,

$$\sum_{\alpha=1}^{n} d_{\kappa\alpha}\bar{d}_{\lambda\alpha} = e_{\kappa\lambda}, \qquad \sum_{\beta=1}^{n} d_{\beta\kappa}\bar{d}_{\beta\lambda} = e_{\kappa\lambda} \qquad (e_{\kappa\kappa} = 1,\ e_{\kappa\lambda} = 0 \text{ für } \kappa \neq \lambda). \tag{3}$$

Mit anderen Worten:

I. *Jede symmetrische Matrix* $C = (c_{\kappa\lambda})$ *laesst sich mit Hilfe einer unitären Substitution auf die Form*

$$UCU' = \begin{pmatrix} \omega_1 & 0 & \cdots & 0 \\ 0 & \omega_2 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & \omega_n \end{pmatrix} \tag{4}$$

bringen, wo $\omega_1, \cdots, \omega_n$ *nichtnegative reelle Zahlen sind.*

Dies bedeutet genauer: Jede symmetrische Bilinearform $f(x, y) = \sum_{\kappa,\lambda=1}^{n} c_{\kappa\lambda} x_\kappa y_\lambda$ laesst sich in der Form

$$f(x, y) = \sum_{\nu=1}^{n} \omega_\nu \phi_\nu(x) \phi_\nu(y) \tag{5}$$

darstellen, wobei die $\phi_\nu(x) = \sum_{\alpha=1}^{n} d_{\nu\alpha} x_\alpha$ den Bedingungen (2) bezw. (3) genügen.

* Received November 8, 1943. This paper has been edited, after the author's death, by Professors Fekete and M. Schiffer of the University of Jerusalem.

1. Bezeichnungen. Jeder Matrix n-ten Grades $C = (c_{\kappa\lambda})$ $(\kappa, \lambda = 1, \cdots, n)$ entspricht einerseits die lineare Substitution (C): $x_\kappa' = \sum_{\lambda=1}^{n} c_{\kappa\lambda} x_\lambda$ $(\kappa = 1, 2, \cdots, n)$ mit unabhängigen Veränderlichen $x_1, \cdots, x_n$, andererseits die Bilinearform

$$C(x, y) = \sum_{\kappa,\lambda=1}^{n} c_{\kappa\lambda} x_\kappa y_\lambda. \tag{6}$$

Für (C) führen wir die Bezeichnung ein: Ist $z = \sum_{\nu=1}^{n} \xi_\nu x_\nu$ eine beliebige Linearform, so bedeutet $z' = z^C$ die Linearform $z' = z^C = \sum_{\nu=1}^{n} \xi_\nu (\sum_{a=1}^{n} c_{\nu a} x_a)$ $= \sum_{a=1}^{n} x_a (\sum_{\nu=1}^{n} c_{\nu a} \xi_\nu)$. Wird (identisch in den x_a)

$$z' = \omega z \quad (\omega \text{ konstant}) \tag{7}$$

und $z \not\equiv 0$, so heisst (bekanntlich) z eine Invariante der Substitution (C). Derartige Invarianten gibt es bekanntlich dann und nur dann, wenn für ω eine charakteristische Wurzel der Gleichung

$$| C - xE | = | c_{\kappa\lambda} - x e_{\kappa\lambda} | = 0$$

gewählt wird.

Zu jeder nicht identisch verschwindenden Linearform z lassen sich n linear unabhängige Linearformen

$$z_\kappa = \sum_{\lambda=1}^{n} s_{\kappa\lambda} x_\lambda, \quad (\kappa = 1, \cdots, n), \qquad | s_{\kappa\lambda} | \neq 0, \quad z_1 = z$$

wählen, und zwar auch solche, die eine unitäre (bezw. reell orthogonale) Substitution bestimmen, wenn z "normiert" ist, *d. h.* $\sum_{\nu=1}^{n} | \xi_\nu |^2 = 1$ ist, bezw. wenn die ξ_ν reell sind und $\sum_{\nu=1}^{n} \xi_\nu^2 = 1$ wird. Bildet man für jedes κ die transformierte Linearform $z'_\kappa = z_\kappa{}^C = \sum_{\lambda=1}^{n} s_{\kappa\lambda} (\sum_{a=1}^{n} c_{\lambda a} x_a)$ und drückt man die z'_κ durch die z_κ aus, so erhält man in (C_1): $z'_\kappa = \sum_{\lambda=1}^{n} c'_{\kappa\lambda} z_\lambda$ $(\kappa = 1, \cdots, n)$ eine zu (C) ähnliche Substitution (C_1), mit der Matrix

$$C_1 = SCS^{-1} \qquad (S = (s_{\kappa\lambda})). \tag{8}$$

Für die Bilinearform $C(x, y)$ gilt ferner folgendes: Ist $S = (s_{\kappa\lambda})$ irgend eine Matrix (Substition) und ersetzt man simultan ("kogredient") die x_κ, y_λ durch $x_\kappa = \sum_{\mu=1}^{n} s_{\mu\kappa} u_\mu$, $y_\lambda = \sum_{\nu=1}^{n} s_{\nu\lambda} v_\nu$, so wird

$$C(x, y) = \sum_{\kappa,\lambda,\mu,\nu} c_{\kappa\lambda} s_{\mu\kappa} s_{\nu\lambda} u_\mu v_\nu = \sum_{\mu,\nu} c_{\mu\nu}'' u_\mu v_\nu$$

wo

$$C_2 = (c_{\mu\nu}'') = SCS'. \tag{9}$$

wird. Hierbei bedeutet S' die zu S "konjugierte," die "transponierte" Matrix. (N. B. Nur dann, wenn S orthogonal ist, wird dies eine Aehnlichkeitstransformation, d. h. $S' = S^{-1}$, nicht dagegen, wenn S nur als unitär, aber komplex bekannt ist, denn alsdann ist nur $S' = \bar{S}^{-1}$.) Im folgenden bezeichne ich in bekannter Weise, wenn C_1, C_2, C_3, C_4 vier Matrizen n-ten Grades $C_a = (c_{\kappa\lambda}{}^{(a)})$ sind, mit $\begin{pmatrix} C_1 & C_2 \\ C_3 & C_4 \end{pmatrix}$ die Matrix des Grades $2n$, die ausführlicher geschrieben, die Gestalt

$$M = \begin{pmatrix} c_{11}{}^{(1)} & \cdots & c_{1n}{}^{(1)} & c_{11}{}^{(2)} & \cdots & c_{1n}{}^{(2)} \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ c_{n1}{}^{(1)} & \cdots & c_{nn}{}^{(1)} & c_{n1}{}^{(2)} & \cdots & c_{nn}{}^{(2)} \\ c_{11}{}^{(3)} & \cdots & c_{1n}{}^{(3)} & c_{11}{}^{(4)} & \cdots & c_{1n}{}^{(4)} \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ c_{n1}{}^{(3)} & \cdots & c_{nn}{}^{(3)} & c_{n1}{}^{(4)} & \cdots & c_{nn}{}^{(4)} \end{pmatrix}$$

hat. Mit diesen "zusammengesetzten" Bildungen kann man im Kalkül additiv und multiplikativ bekanntlich ebenso operieren wie mit Zahlenmatrizen, nur dass man bei der Multiplikation auf die Reihenfolge der Faktoren achtgeben muss. Man beachte insbesondere, dass die Transponierte von M gleich $M' = \begin{pmatrix} C_1' & C_3' \\ C_2' & C_4' \end{pmatrix}$ zu setzen ist.

2. Der Stickelbergersche Beweis für das Hauptachsentheorem. Um das Folgende verständlicher zu machen, gebe ich hier den schönen, klassisch gewordenen, Stickelbergerschen Beweis für das Hauptachsentheorem wieder. Ist $C = (c_{\kappa\lambda})$ eine *reelle* symmetrische Matrix und ω eine charakteristische Wurzel von C, so bestimme man eine zugehörige invariante Linearform $z = \sum_{\nu=1}^{n} \xi_\nu x_\nu$ mit $z^C = \omega z$, was auf

$$\sum_{\nu=1}^{n} c_{\nu a} \xi_\nu = \omega \xi_a \tag{10}$$

führt. Die $\xi_1, \cdots, \xi_n$ brauchen zunächst noch nicht reell zu sein, man darf aber den noch unbestimmt bleibenden Proportionalitätsfaktor so wählen,

dass $|\xi_1|^2+|\xi_2|^2+\cdots+|\xi_n|^2=1$ wird. Aus (10) folgt aber (Cauchy!) wegen $c_{\nu\alpha}=c_{\alpha\nu}=\bar{c}_{\nu\alpha}$

$$\omega=\omega\sum_{\alpha=1}^{n}|\xi_\alpha|^2=\sum_{\nu,\alpha}c_{\nu\alpha}\xi_\nu\bar{\xi}_\alpha=\sum_{\nu,\alpha}c_{\alpha\nu}\xi_\nu\bar{\xi}_\alpha$$

und

$$\bar{\omega}=\sum_{\nu,\alpha}c_{\nu\alpha}\bar{\xi}_\nu\xi_\alpha=\sum_{\nu,\alpha}c_{\alpha\nu}\bar{\xi}_\alpha\xi_\nu=\omega;$$

also ist ω reell, und folglich darf man $\xi_1,\cdots,\xi_n$ auch reell wählen. Wählt man, was jedenfalls möglich ist, n Linearformen $z_\kappa=\sum_{\lambda=1}^{n}s_{\kappa\lambda}x_\lambda$, die eine reelle orthogonale Transformation bestimmen, so dass $z_1=z$ wird, so führt die Einführung von $z_1,z_2,\cdots,z_n$ als neue Veränderliche auf eine Aehnlichkeitstransformation (vgl. 1) $SCS^{-1}=(c'_{\kappa\lambda})$, worin wegen $z'=\omega z$, $z=z_1$ in der ersten Zeile $c'_{11}=\omega_1$, $c'_{12}=\cdots=c'_{1n}=0$ wird. Da aber $S^{-1}=S'$ wird, erhalten wir in SCS^{-1} auch SCS', d. h. von neuem Symmetrie, also $c'_{\kappa\lambda}=c'_{\lambda\kappa}$;

insbesondere wird $SCS^{-1}=\begin{pmatrix}\omega & 0 & \cdots & 0\\ 0 & d_{22} & \cdots & d_{2n}\\ \cdot & \cdot & & \cdot\\ \cdot & \cdot & & \cdot\\ \cdot & \cdot & & \cdot\\ 0 & d_{n2} & \cdots & d_{nn}\end{pmatrix}$. Die symmetrische reelle Matrix $(d_{\nu\mu})$ kann man nun weiterhin ebenso behandeln usw.

3. Der Haupthilfssatz. Ist $U=(u_{\kappa\lambda})$ eine unitäre Substitution n-ten Grades, so setze man

$$u_{\kappa\lambda}=p_{\kappa\lambda}+iq_{\kappa\lambda},\qquad (p_{\kappa\lambda})=P,\quad (q_{\kappa\lambda})=Q.$$

Jede symmetrische Matrix $C=(c_{\kappa\lambda})$ mit $c_{\kappa\lambda}=a_{\kappa\lambda}+ib_{\kappa\lambda}$, $(a_{\kappa\lambda})=A$, $(b_{\kappa\lambda})=B$ liefert dann bei der Transformation UCU' eine neue Zerlegung in Real- und Imaginärteil $UCU'=A^*+iB^*$. Hierbei wird

$$UCU'=(P+iQ)(A+iB)(P'+iQ')=A^*+iB^*,$$

woraus folgt

$$\begin{aligned}A^*&=PAP'-QBP'-PBQ'-QAQ'\\ B^*&=PAQ'-QBQ'+PBP'+QAP'.\end{aligned}\tag{11}$$

Operiert man ferner bei der unitären Substitution U mit komplexen Veränderlichen $x_\nu=u_\nu+iv_\nu$, so liefert

$$x'_\kappa=u'_\kappa+iv'_\kappa=\sum_{\lambda=1}^{n}u_{\kappa\lambda}x_\lambda=\sum_{\lambda=1}^{n}(p_{\kappa\lambda}+iq_{\kappa\lambda})(u_\lambda+iv_\lambda)$$

die Formeln

$$(12) \qquad u'_\kappa = \sum_{\lambda=1}^{n} (p_{\kappa\lambda} u_\lambda - q_{\kappa\lambda} v_\lambda), \qquad v'_\kappa = \sum_{\lambda=1}^{n} (q_{\kappa\lambda} u_\lambda + p_{\kappa\lambda} v_\lambda).$$

Hier handelt es sich um eine lineare Transformation in den $2n$ reellen Veränderlichen, $u_1, \cdots, u_n, v_1, \cdots, v_n$ die wegen $\sum_{\kappa=1}^{n} |x'_\kappa|^2 = \sum_{\kappa=1}^{n} |x_\kappa|^2$ auf

$$u'_1{}^2 + \cdots + u'_n{}^2 + v'_1{}^2 + \cdots + v'_n{}^2 = u_1{}^2 + \cdots + u_n{}^2 + v_1{}^2 + \cdots + v_n{}^2,$$

also auf eine relle orthogonale Transformation führt. *Hält man an der Reihenfolge* $u_1, \cdots, u_n, v_1, \cdots, v_n$ *fest,* so erscheint als Matrix der Transformation die Matrix

$$(12') \qquad D = \begin{pmatrix} P & -Q \\ Q & P \end{pmatrix}.$$

Man bilde nun mit Hilfe der Matrix $C = A + iB$ die "zugeordnete" *reelle symmetrische Matrix des Grades* $2n$ $R = \begin{pmatrix} A & B \\ B & -A \end{pmatrix} = R(C) = R(A + iB)$. Es gilt dann der für uns ausschlaggebende Hilfssatz:

II. *Ist* U *eine unitäre Transformation* $U = P + iQ$ *in den* n *Veränderlichen* $x_1, \cdots, x_n$, *und bildet man neben* $C = A + iB$ *noch* $UCU' = A^* + iB^*$, *so hat die Bildung* $R(C)$ *die Grundeigenschaft:*

$$(13) \qquad DR(C)D' = R(UCU') = \begin{pmatrix} A^* & B^* \\ B^* & -A^* \end{pmatrix}.$$

Der Beweis beruht auf einer einfachen Rechnung: Es wird

$$DR(C)D' = \begin{pmatrix} P & -Q \\ Q & P \end{pmatrix} \begin{pmatrix} A & B \\ B & -A \end{pmatrix} \begin{pmatrix} P' & Q' \\ -Q' & P' \end{pmatrix} = \begin{pmatrix} PA - QB, PB + QA \\ QA + PB, QB - PA \end{pmatrix} \begin{pmatrix} P' \\ -Q' \end{pmatrix}$$

$$= \begin{pmatrix} PAP' - QBP' - PBQ' - QAQ', \; PAQ' - QBQ' + PBP' + QAP' \\ QAP' + PBP' - QBQ' + PAQ', \; QAQ' + PBQ' + QBP' - PAP' \end{pmatrix}.$$

Ein Vergleich mit den Formeln (11) liefert unseren Hilfssatz. Man beachte noch folgendes: Die Schreibweise $R = \begin{pmatrix} A & B \\ B & -A \end{pmatrix}$ ist an die Reihenfolge $u_1, \cdots, u_n, \; v_1 \cdots v_n$ geknüpft. Zieht man die natürliche Reihenfolge $u_1, v_1, \cdots, u_n, v_n$ vor, so erhält die reelle symmetrische Matrix R die Koeffizientenmatrix $R_1 = \begin{pmatrix} a_{11} & b_{11} & a_{12} & b_{12} \cdots \\ b_{11} & -a_{11} & b_{12} & -a_{12} \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}$. Hierbei geht R_1 aus R durch eine Aehnlichkeitstransformation hervor. Man beachte noch folgendes: Die charakteristischen Wurzeln einer Matrix der Form R haben (auch für

beliebige zwei Teilmatrizen A und B) die Gestalt $\omega_1, \omega_2, \cdots, \omega_n, -\omega_1, -\omega_2, \cdots, -\omega_n$. Dies folgt am einfachsten aus der Formel

$$\begin{pmatrix} 0 & E \\ -E & 0 \end{pmatrix} \begin{pmatrix} A & B \\ B & -A \end{pmatrix} = \begin{pmatrix} -A & -B \\ -B & A \end{pmatrix} \begin{pmatrix} 0 & E \\ -E & 0 \end{pmatrix},$$

woraus hervorgeht, dass R und $-R$ einander aehnliche Matrizen sind, dass also die charakteristische Determinante von R eine gerade Funktion der Unbekannten ist. Hieraus folgt, dass R jedenfalls im symmetrischen Fall auch nichtnegative charakteristische Wurzeln besitzt.

4. Beweis des Satzes I. Für unsere symmetrische Matrix $C = (c_{\kappa\lambda})$, $c_{\kappa\lambda} = a_{\kappa\lambda} + ib_{\kappa\lambda}$ suchen wir diejenigen nicht identisch verschwindenden Linearformen $\psi = \sum_{\nu=1}^{n} \xi_\nu x_\nu$, für die die transformierte Funktion $\psi' = \sum_{\nu=1}^{n} \xi_\nu \sum_{a=1}^{n} c_{\nu a} x_a$ sich nur um einen konstanten Faktor $\tilde{\omega}$ von $\overline{\psi} = \sum_{\nu=1}^{n} \overline{\xi}_\nu x_\nu$ unterscheidet, d. h. es soll $\psi' = \psi^C = \tilde{\omega}\overline{\psi}$ sein. Hier kann nur der absolute Betrag von $\tilde{\omega}$ bestimmt sein. Denn existiert eine solche Funktion und setzt man für alle $\xi_\nu = e^{i\phi}\xi_\nu^{(1)}$, so erhält man $e^{i\phi}(\sum_{\nu=1}^{n} \xi_\nu^{(1)} x_\nu)' = \tilde{\omega}e^{-i\phi} \sum_{\nu=1}^{n} \overline{\xi_\nu^{(1)}} x_\nu$. Ist aber $|\tilde{\omega}| = \omega$, $\tilde{\omega} = \omega e^{i\alpha}$, so setze man $\phi = \alpha/2$. An Stelle von $\tilde{\omega}$ tritt dann in $\psi' = \tilde{\omega}\psi$ die reelle nichtnegative Grösse ω. Man betrachte also von vornherein nur die "Halbinvarianten" $\psi = \sum_{\nu=1}^{n} \xi_\nu x_\nu$, für die bei reellem nicht negativem ω, $\psi' = \omega\psi$ oder, was dasselbe ist,

$$\sum_{\nu=1}^{n} c_{\nu a}\xi_\nu = \omega\xi_a \tag{14}$$

wird. Setzt man $\xi_\nu = \zeta_\nu - i\eta_\nu$, so erhält man

$$\sum_{\nu=1}^{n} (a_{\nu a} + ib_{\nu a})(\zeta_\nu - i\eta_\nu) = \omega(\zeta_a + i\eta_a)$$

also

$$\sum_{\nu=1}^{n} (a_{\nu a}\zeta_\nu + b_{\nu a}\eta_\nu) = \omega\zeta_a; \qquad \sum_{\nu=1}^{n} (b_{\nu a}\zeta_\nu - a_{\nu a}\eta_\nu) = \omega\eta_a. \tag{15}$$

Dies besagt aber nur, dass im symmetrischen Fall C Halbinvarianten mit reellem nichtnegativem ω bezitzt; es gehört nämlich zu jeder solchen „Wurzel" der zugeordneten Matrix $R = R(C)$ eine solche Halbinvariante von C. Genauer: Jede Halbinvariante $\psi = \sum_{\nu=1}^{n} (\zeta_\nu - i\eta_\nu)x_\nu$ von C bestimmt in $\Psi(u, v) = \sum_{\nu=1}^{n} \zeta_\nu u_\nu$

$+\sum_{\nu=1}^{n}\eta_\nu v_\nu$ eine zum gleichen Faktor gehörende Invariante der reellen Matrix $R=R(C)$. Um nun den Satz I zu beweisen, schliesse man so: Man bestimme eine Wurzel $\omega \geq 0$ von $R=R(C)$ und eine zugehörige Halbinvariante $\psi=\sum_{\nu=1}^{n}\xi_\nu x_\nu$ von C. Diese ξ_ν darf man als normiert ansehen, da man sonst die ξ_ν durch $\frac{\xi_\nu}{\sqrt{|\xi_1|^2+\cdots+|\xi_n|^2}}$ ersetzen kann, was die Forderung an ψ nicht beeinflusst. Die Linearform ψ ergänze man zu einem System $\psi=\psi_1,\psi_2,\cdots,\psi_n$ von Linearformen derart, dass $|\psi_1|^2+\cdots+|\psi_n|^2 = |x_1|^2+\cdots+|x_n|^2$ wird. Sie liefern dann ein System mit unitärer Matrix $U=P+iQ$. In den reellen Veränderlichen u_ν, v_ν liefern sie $2n$ reelle Linearformen $\Psi_1,\cdots,\Psi_{2n}$ mit $\Psi_1=\Psi(u,v)$, die ein Orthogonalsystem mit der Koeffizientenmatrix $D=\begin{pmatrix}P & -Q\\ Q & P\end{pmatrix}$ bestimmen. Bildet man (vgl. **2**) DRD^{-1} so enthält diese Matrix, weil $\Psi_1=\Psi(u,v)$ eine zu ω gehörende Invariante ist, in der ersten Zeile nur die Grössen $\omega, 0, 0,\cdots, 0$ und wegen der Symmetrie von R (und wegen $D^{-1}=D'$) gilt dasselbe auch für die erste Kolonne. Der Hilfssatz II lehrt uns aber, dass, wenn $UCU'=A^*+iB^*$ gesetzt wird, die erste Zeile von $DRD'=DRD^{-1}$ nur die Elemente der ersten Zeilen von A^* und B^* enthält. Hieraus folgt aber, dass die von uns gewonnene Matrix UCU' die Gestalt $UCU'=\begin{pmatrix}\omega & 0 & \cdots & 0\\ 0 & c^*_{22} & \cdots & c^*_{2n}\\ \cdot & \cdot & & \cdot\\ \cdot & \cdot & & \cdot\\ \cdot & \cdot & & \cdot\\ 0 & c^*_{n2} & \cdots & c^*_{nn}\end{pmatrix}$ hat. Das genügt aber, um den Satz I aussprechen zu können. Denn die Matrix der $c^*_{\kappa\lambda}$ $(\kappa,\lambda=2,3\cdots n)$ ist wieder symmetrisch und kann in derselben Weise weiter behandelt werden, was zuletzt auf eine Schlussform $WCW'=\begin{pmatrix}\omega_1 & 0 & \cdots & 0\\ 0 & \omega_2 & & \\ \cdot & & \cdot & \\ \cdot & & & \cdot\\ \cdot & & & \cdot\\ 0 & \cdots & \cdots & \omega_n\end{pmatrix}$ mit nichtnegativem reellen ω_ν führt, bei unitärem W.

N. B. Hier ergibt sich, dass die ω eindeutig als die sämtlichen p positiven Wurzeln von $R=R(C)$ charakterisiert werden koennen, wozu noch $n-p$ Nullen hinzukommen. Man kann die $\omega_\nu{}^2$ aber auch als die Wurzeln der beiden "Normen" $\bar{C}C$ *und* $C\bar{C}$, die nichtnegative Hermitesche Formem liefern, kennzeichnen. Dies ergibt sich aus

$$C_1 = WCW',\ C_1\bar{C}_1 = WCW'\bar{W}\bar{C}\bar{W}' = WC\bar{C}W^{-1} = \begin{pmatrix} \omega_1^2 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \omega_2^2 & \cdot & \cdot & \cdot & 0 \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ 0 & \cdot & \cdot & \cdot & \cdot & \omega_n^2 \end{pmatrix};$$

analog für $\bar{C}C$ (vgl. die Bemerkung am Schluss der Notiz).

5. Folgerungen aus I. Wir wissen nunmehr, dass jede symmetrische Bilinearform $f = \sum_{\kappa=1}^{n} c_{\kappa\lambda} x_\kappa y_\lambda$ sich in der folgenden Form schreiben laesst:

$$f(x,y) = \sum_{\kappa,\lambda=1}^{n} c_{\kappa\lambda} x_\kappa y_\lambda = \omega_1 \phi_1(x)\,\phi_1(y) + \cdots + \omega_n \phi_n(x)\,\phi_n(y)$$

mit

$$\sum_{\nu=1}^{n} |\,\phi_\nu(x)\,|^2 = \sum_{\nu=1}^{n} |\,x_\nu\,|^2, \qquad \sum_{\nu=1}^{n} |\,\phi_\nu(y)\,|^2 = \sum_{\nu=1}^{n} |\,y_\nu\,|^2.$$

Die $\phi_1, \cdots, \phi_n$ kann man wegen der linearen Unabhängigkeit eines unitären Systems beliebig wählen, also auch so, dass $\phi_\nu = 1$, $\phi_\lambda = 0$ für $\lambda \neq \nu$ wird. Folglich sind die Werte $\omega_1, \omega_2, \cdots, \omega_n$ Werte der *quadratischen Form* $f(x,x)$ mit $|\,x_1\,|^2 + \cdots + |\,x_\nu\,|^2 = 1$. Für $\sum_{\nu=1}^{n} |\,x_\nu\,|^2 \leqq 1$, $\sum_{\nu=1}^{n} |\,y_\nu\,|^2 \leqq 1$ erhalten wir ferner, wenn $\omega_1 \geqq \omega_2 \geqq \cdots \geqq \omega_n$ ist, nach der Schwarzschen Ungleichung

$$|\,f(x,y)\,| \leq \omega_1 \sum_{\nu=1}^{n} |\,\phi_\nu(x)\,|\;|\,\phi_\nu(y)\,|$$

$$\leq \omega_1 \sqrt{\sum_{\nu=1}^{n} |\,\phi_\nu(x)\,|^2 \sum_{\nu=1}^{n} |\,\phi_\nu(y)\,|^2} = \omega_1 \sqrt{\sum_{\nu=1}^{n} |\,x_\nu\,|^2 \sum_{\nu=1}^{n} |\,y_\nu\,|^2} \leq \omega_1.$$

Das gibt den für reelle $c_{\kappa\lambda}$ und x_ν, y_ν bekannten Satz:

III. *Jede symmetrische Bilinearform $f(x,y)$ hat die Eigenschaft, dass im Hilbertschen Raume $\sum_{\nu=1}^{n} |\,x_\nu\,|^2 \leq 1$, $\sum_{\nu=1}^{n} |\,y_\nu\,|^2 \leq 1$ das Maximum ihres absoluten Betrages auf dem Rande, und zwar schon für $x_1 = y_1, \cdots, x_n = y_n$, d.h. von der zugehörigen quadratischen Form $f(x,x)$ angenommen wird.*

Als unmittelbare Anwendungen ergeben sich Bemerkungen über quadratische Formen $\sum_{\kappa,\lambda=1}^{\infty} c_{\kappa\lambda} x_\kappa x_\lambda$ $(c_{\kappa\lambda} = c_{\lambda\kappa})$ mit unendlich vielen Veraenderlichen, die im Hilbertschen Raume $\sum_{\nu=1}^{\infty} |\,x_\nu\,|^2 \leq 1$ beschränkt sein sollen:

IV. *Eine quadratische Form $\sum_{\kappa,\lambda=1}^{\infty} c_{\kappa\lambda} x_\kappa x_\lambda$ ist bei $c_{\kappa\lambda} = c_{\lambda\kappa}$ auch für komplexe $c_{\kappa\lambda}$ dann und nur dann beschränkt, wenn dies für die zugehörige*

Bilinearform der Fall ist. Quadratische Form und Bilinearform besitzen ferner die gleiche obere Schranke.

Es kommt weiter hinzu: Setzt man $c_{\kappa\lambda} = a_{\kappa\lambda} + ib_{\kappa\lambda}$ so ist die komplexe quadratische Form dann und nur dann beschränkt, wenn dies für die reelle quadratische Form $\sum_{\kappa,\lambda=1}^{\infty} [a_{\kappa\lambda}(u_\kappa u_\lambda - v_\kappa v_\lambda) + b_{\kappa\lambda}(u_\kappa v_\lambda + v_\kappa u_\lambda)]$ in den Veränderlichen $u_1, v_1, u_2, v_2 \cdots$ gilt. Beim Uebergang von dem komplexen Fall zum reellen aendert sich die obere Schranke nicht.

Es ist also für $c_{\kappa\lambda} = c_{\lambda\kappa}$ nicht erforderlich, die "Normen" $C\bar{C}'$ bezw. $\bar{C}'C$ einzuführen.

Zusatz. Der Zusammenhang zwischen $R = \begin{pmatrix} A & B \\ B & -A \end{pmatrix}$ und den Normen

$$C\bar{C} = (A + iB)(A - iB)$$
$$= A^2 + B^2 + i(BA - AB) \text{ bezw. } \bar{C}C = A^2 + B^2 - i(BA - AB)$$

wird klar, wenn man $R^2 = \begin{pmatrix} A^2 + B^2, & AB - BA \\ BA - AB, & A^2 + B^2 \end{pmatrix}$ bildet. Diese Matrix hat die Gestalt $\begin{pmatrix} F & G \\ -G & F \end{pmatrix}$. Es wird aber

$$\begin{pmatrix} E & -iE \\ -iE & E \end{pmatrix} \begin{pmatrix} F & G \\ -G & F \end{pmatrix} = \begin{pmatrix} F + iG & 0 \\ 0 & F - iG \end{pmatrix} \begin{pmatrix} E & -iE \\ -iE & E \end{pmatrix}$$

wobei

$$\begin{pmatrix} E & -iE \\ -iE & E \end{pmatrix} \begin{pmatrix} E & iE \\ iE & E \end{pmatrix} = \begin{pmatrix} 2E & 0 \\ 0 & 2E \end{pmatrix}$$

wird. Dies zeigt, dass R^2 der Matrix $\begin{pmatrix} \bar{C}C & 0 \\ 0 & C\bar{C} \end{pmatrix}$ ähnlich ist. Die charakteristischen Wurzeln von $C\bar{C}$ (bezw. $\bar{C}C$) sind also in der Tat die Quadrate der Wurzeln von R, von zwei gleichen Quadraten nur eines gerechnet.

MEAN-VALUES OF ARITHMETICAL REPRESENTATIONS.*

By AUREL WINTNER.

By a representation, $f = f(n)$, will be meant any (number-valued) multiplicative representation of the ordinary multiplication on the semi-group of all positive integers; that is, any sequence of numbers $f(1), f(2), \cdots$ satisfying the relation $f(nm) = f(n)f(m)$ for every pair of positive integers n, m, distinct or not. Thus, if the trivial case $f(1) = 0$ (which leads to the identical vanishing of f) is excluded, then $f(1)$ must be 1, and so $f(n)$ becomes uniquely determined for every n by an arbitrary assignment of the values $f(p)$, where p runs through all primes.

It was recently shown ([2], p. 46) that every representation $f(n)$ which attains only the values ± 1 must have a mean

$$M(f) = \lim_{n\to\infty} \frac{f(1) + \cdots + f(n)}{n} . \tag{1}$$

The proof depended on *the prime number theorem* (which, in turn, follows by choosing $f(n) = \lambda(n)$, where λ denotes Liouville's function). In the present note, a parallel, but *elementary*, theorem will be verified by showing that every representation $f(n)$ which is bounded and non-negative has a mean (1). Here, and in the sequel, "non-negative" means "real and non-negative".

A class of functions more general than that of the representations results if the condition $f(nm) = f(n)f(m)$ is required only for integers n, m which are relatively prime. For these functions $f(n)$, the so-called multiplicative functions, not only the values $f(p)$, attained at primes, but all the values $f(p), f(p^2), \cdots$, attained at prime powers, can be chosen arbitrarily. The representations result if $f(p^k)$ is chosen to be the k-th power of the value $f(p)$, where p and k run through all primes and all positive integers respectively.

The theorem, quoted above concerning the case $f(n) = \pm 1$ of representations, was derived in [2], p. 46, in its more general form:

(i) *Every multiplicative function $f(n)$ which attains only the values ± 1 has a mean $M(f)$.*

* Received March 22, 1945.

Correspondingly, the elementary fact, stated above concerning representations, will be proved for the general case also:

(ii) *Every multiplicative function $f(n)$ which is non-negative and bounded has a mean $M(f)$.*

If $f(n) = \phi(n)/n$ or $f(n) = |\mu(n)|$, where ϕ and μ denote Euler's and Möbius' functions respectively, it is clear that the assumptions of (ii) are satisfied. Thus (ii) supplies the existence of (1) in both cases. In the first case, the existence of (1) goes back to Dirichlet (in fact, it is contained in his result concerning the sum-function of ϕ). In the second case, the existence of (1) means that the sequence of the square-free numbers has a density.

In both of these classical cases, the mere existence of (1) can of course be improved by *explicit* estimates of the error belonging to a fixed n. But (ii) is a *general* fact and cannot, therefore, supply improved remainder terms. For instance, (ii) contains the fact (cf., e. g., [1], p. 68) that every multiplicative set of positive integers has a density. But this particular case of (ii) is needed for the extension of the prime number theorem to the case of arbitrary sets of primes ([1], p. 71). In addition, this particular case of (ii) contains the truth of a conjecture of Ramanujan ([2], p. 31).

All of this seems to warrant an isolation of the general fact expressed by (ii).

In the proof, use will be made of the following elementary criterion: If $g(n)$ is a multiplicative function satisfying

$$\sum_p \sum_{k=1}^{\infty} |g(p^k) - g(p^{k-1})|/p^k < \infty \tag{2}$$

(where $g(p^{k-1})$ is $g(1) = 1$ when $k = 1$), then the mean $M(g)$ exists; cf. [1], p. 41. It is also known that, if (2) is satisfied, the mean $M(g)$ of the multiplicative function g can be represented by the Eulerian product

$$M(g) = \prod_p \left(1 + \sum_{k=1}^{\infty} \{g(p^k) - g(p^{k-1})\}/p^k\right); \tag{3}$$

cf. [1], p. 14.

What will actually be shown is that, if a multiplicative function $f(n)$ satisfies the assumptions of (ii) but is such that (2) fails to be satisfied for $g = f$, then $M(f)$ must exist and be 0. But it is known (cf. [1], p. 41) that, if a multiplicative function g satisfies (2), then g must be almost-periodic (B). On the other hand, if $M(f)$ exists and is 0, then, since $f \geqq 0$ by assumption, it is clear that f is almost-periodic (B), with $0 + 0 \cdots$ as Fourier series (B). Consequently, (ii) can be refined as follows:

(ii bis) *Every multiplicative function which is non-negative and bounded is almost-periodic* (B).

The possibility of this refinement of (ii) is due entirely to the *arithmetical* assumption of multiplicativity. In other words, if a function $f(n)$ is non-negative and bounded, then the mere existence of $M(f)$ does not imply the almost-periodicity (B) of $f(n)$. For a simple example proving more than the falsity of this implication, cf. [1], p. 28.

Let $f(n)$ be a function satisfying the assumptions of (ii). Thus

$$0 \leqq f(n) < A, \tag{4}$$

where A is a constant. In particular $|f(p^k)| < A$. Since $f(p^0) = f(1)$ is 1, it follows that (2) is satisfied for $g = f$ unless

$$\sum_p |f(p) - 1|/p = \infty. \tag{5}$$

Consequently, both (ii) and (ii bis) will be proved if it is shown that, if f satisfies the assumptions of (ii) and the additional assumption (5), then $M(f)$ exists and is 0.

To this end, let the sequence of all primes p be split into two complementary subsequences (either of which may be finite or vacuous), as follows: A prime p is in the first or in the second subsequence according as the value $f(p)$ is or is not less than 1. Thus, if a prime p is called a prime q or a prime r according as it is in the first or in the second subsequence,

$$(6_1) \qquad 0 \leqq f(q) < 1; \qquad\qquad (6_2) \qquad 1 \leqq f(r).$$

It is easy to see that

$$\sum_q |f(q) - 1|/q = \infty. \tag{7}$$

In fact, since q and r are complementary, it is clear from the assumption (5) that more than (7) is contained in the assertion

$$\sum_r |f(r) - 1| < \infty. \tag{8}$$

But (8) can be verified as follows:

Let $r_1, r_2, \cdots, r_j, \cdots$ denote the sequence of all primes r. It may be assumed that this sequence is infinite, since (8) is clear in the contrary case. For a fixed j, choose $n = r_1 r_2 \cdots r_j$ in (4). Then, since f is a multiplicative function, it follows that the partial products of the infinite product $\Pi f(r)$ form a bounded sequence. On the other hand, (6_2) shows that this sequence

is non-decreasing. Accordingly, the infinite product $\Pi f(r)$ is convergent. In view of (6_2), this is equivalent to (8).

With reference to the given function $f(n)$ and to a fixed positive number t, let $f_t(n)$ denote the multiplicative function defined by the data $f_t(p^k)$ which are assigned as follows: If either p is an r or k is greater than 1, then $f_t(p^k) = f(p^k)$; while in the remaining case (that is, if p is a prime q and k is 1) the value of $f_t(p^k) = f_t(q)$ is $f(q)$ or 1 according as q is or is not less than t. Thus it is clear from (4) that

$$0 \leqq f_t(p^k) < A, \tag{9}$$

since A is greater than 1. Furthermore, it is seen from (6_1) and from the first of the inequalities (4) that the inequalities

$$0 \leqq f(n) \leqq f_t(n) \tag{10}$$

hold for every prime power, $n = p^k$. However, since both functions f, f_t are multiplicative, (10) must hold without the restriction $n = p^k$ also.

Condition (2) is satisfied by the function $g = f_t$ (for every fixed t). In fact, it is clear from (9) that the convergence of the double series (2) in the case $g = f_t$ is equivalent to the convergence of the simple series belonging to $k = 1$. Since $f_t(1) = 1$, it follows that (2) is equivalent to the convergence of

$$\sum_p |f_t(p) - 1|/p = \sum_r |f_t(r) - 1|/r + \sum_q |f_t(q) - 1|/q.$$

But the convergence of the first series on the right is implied by (8), while the second series on the right has just a finite number of non-vanishing terms, since $f_t(q)$ was chosen to be 1 for every q exceeding t.

Since (2) is satisfied by $g = f_t$, it follows that $M(f_t)$ exists and is represented by the case $g = f_t$ of the absolutely convergent product (3). In view of the definition of $f_t(p^k)$, those factors of this infinite product which belong to primes r are independent of t. Hence, the product contributed by all these factors together can be denoted by C (without a subscript t). Then $M(f_t) = CQ_t$, where Q_t is an abbreviation for the product of all factors which belong to primes q.

Let Q_t be split into G_t times H_t, hence $M(f_t)$ into

$$M(f_t) = CG_tH_t, \tag{11}$$

where G_t comprises those factors of the product Q_t which belong to primes q having a value less than t, and H_t represents the contribution of all the remaining primes q. It is seen from the definition of $f_t(p^k)$ that, whether q

belongs to G_t or to H_t, the difference $f_t(q^k) - f_t(q^{k-1})$ is identical with $f(q^k) - f(q^{k-1})$ except when $k = 1$; and that, in the exceptional case $k = 1$, the difference $f_t(q^1) - f_t(q^0) = f_t(q) - 1$ is $f(q) - 1$ or $1 - 1 = 0$ according as q belongs to G_t or to H_t. If these data are substituted into the definitions of the contributions G_t, H_t to the case $g = f_t$ of the product (3), it follows that

$$G_t = \prod_{q<t} (1 + \{f(q) - 1\}/q + \sum_{k=2}^{\infty} \{f(q^k) - f(q^{k-1})\}/q^k) \tag{12}$$

and

$$H_t = \prod_{q \geqq t} (1 + \sum_{k=2}^{\infty} \{f(q^k) - f(q^{k-1})\}/q^k). \tag{13}$$

According to (12) and (4),

$$G_t = \prod_{q<t} (1 - \{1 - f(q)\}/q + O(q^{-2})).$$

It follows therefore from (6_1) that G_t is majorized by a constant multiple of

$$\prod_{q<t} (1 - \{1 - f(q)\}/q).$$

In view of (7) and (6_1), this means that $G_t \to 0$ as $t \to \infty$. Hence, in order to conclude from (11) that

$$M(f_t) \to 0 \quad \text{as} \quad t \to \infty, \tag{14}$$

it is sufficient to observe that, according to (13) and (4), the function H_t of t is majorized by the constant

$$\prod_{q} (1 + \sum_{k=2}^{\infty} 2A/q^k) < \infty.$$

Finally, it is clear from (14) and (10) that the limit (1) exists and is 0. But this is precisely what had to be proved.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] A. Wintner, *Eratosthenian Averages*, Baltimore, 1943.
[2] ———, *The Theory of Measure in Arithmetical Semi-Groups*, Baltimore, 1944.

DIRICHLET SERIES MEROMORPHIC IN A HALF-PLANE.*

By H. D. Brunk.

1. Introduction and basic theorems. The usefulness of Hadamard's famous "multiplication" theorem for Taylor series is well-recognized. Its direct analogue for Dirichlet series is very easily shown not to hold. However, Mandelbrojt has given powerful theorems ([4] and [5]; see also [6]) which generalize Hadamard's theorem for Dirichlet series. Generalizations of one of Mandelbrojt's theorems have been discussed in a recent paper of the author [2], and applications of these theorems and ideas form the subject matter of the present paper.

A function $f(s)$ is said to be represented by a Dirichlet series if

$$f(s) = \sum a_n e^{-\lambda_n s}, \tag{1}$$

where s is a complex variable, $s = \sigma + it$, $\{a_n\}$ is a sequence of complex constants, and $\{\lambda_n\}$ is a sequence of non-negative real numbers, strictly increasing to infinity. If the sequence $\{\lambda_n\}$ is the sequence of non-negative integers, the series is called a Taylor-Dirichlet series. Equation (1) is to be understood as implying that $f(s)$ is given by the series where the series converges, and is defined by analytic continuation wherever possible. Only uniform functions will be considered; this restriction is to be understood, whether or not it is specifically stated. However, the theorems of the paper apply to non-uniform functions as well, if cuts which render them uniform are allowed to play the rôles otherwise assigned to singular points of uniform functions. Also, only those functions will be considered here which are represented by Dirichlet series having finite abscissae of absolute convergence. V. Bernstein's notation will be used: $\mathcal{A}_f$ denotes the abscissa of absolute convergence of the function $f(s)$, and $\mathcal{H}_f$ denotes its abscissa of holomorphism.

The definitions of the next few paragraphs will be required in order to state the fundamental theorem, Theorem 1. The symbol $+$, as an operator on sets of points, will be used as it is commonly used in point set theory; so that if A and B are sets, $A + B$ is their sum, or union. Also $A \cdot B$ denotes their intersection, the set of all points common to both. The word curve, when

* Received August 27, 1944; Presented to the American Mathematical Society April 28, 1944.

used, refers to a simple Jordan curve. The term domain will refer to an open, connected set of points; the term region will be used in referring to a domain plus all, some, or none of its boundary points. A point will be considered singular for a function if it is not possible to continue the function analytically throughout a complete neighborhood of the point.

DEFINITION. A domain O is *bounded connectedly with respect to a domain* D if there exists a positive number K such that to every point of O corresponds a simple Jordan curve in $O + D$ of length less than K which joins this point to points in D. It is of course implied that $O + D$ is itself a domain.

Let $f(s)$ be given by (1), and let S_f denote the set of singular points of $f(s)$ in the complex s-plane, where $s = \sigma + it$. Let c be a real number greater than α_f, and let A be a closed set in $\sigma < c$ containing S_f.

DEFINITION. *The function* $f(s)$ *is "* M *" in* $\sigma > \sigma_1$ *with respect to* A *if* $f(s)$ *is bounded in every domain which is bounded connectedly with respect to* $\sigma > c$ *and whose boundary is at positive distance from the set* $A + P_1$, *where* P_1 *is the half-plane* $\sigma \leq \sigma_1$. Since $f(s)$ is bounded in the half-plane $\sigma \geq c$ if $c > \alpha_f$, the definition is clearly independent of the choice of the real number c greater than α_f. *If* $f(s)$ *satisfies the above condition for every real number* σ_1, *then* $f(s)$ *is "* M *" in the entire plane with respect to* A. *If* $f(s)$ *satisfies the condition with* A *replaced by* S_f, *then* $f(s)$ *is "* M *" in* $\sigma > \sigma_1$.

The function $f(s)$ is "*bounded except for singularities*" in $\sigma > \sigma_1$ if $f(s)$ is bounded in that part of the half-plane $\sigma > \sigma_1$ remaining after circles of arbitrary, fixed radius about points of S_f have been extracted. It has been observed [2] that *if a function* $f(s)$ *is bounded except for singularities in a certain half-plane, it is also "*M*" in that half-plane.* The converse is not true. It has also been shown [2] that *if* $f(s)$ *fails to take two distinct values* a *and* b *in a strip* $\sigma_1 < \sigma < c$ *for real numbers* σ_1 *and* c, *where* $c > \alpha_f$, *then* $f(s)$ *is "*M*" in* $\sigma > \sigma_1$. In fact, *this is true if* $f(s)$ *takes at most a finite number of times in* $\sigma_1 < \sigma < c$ *each of two values* a *and* b. For it was seen (Lemma 1 of [2]) that $f(s)$ is bounded in every domain bounded connectedly with respect to $\sigma > c$ which is at positive distance from $A + P_1$, where $A = S_f + E_f(a) + E_f(b)$, and $P_1 = (\sigma \leq \sigma_1)$; here $E_f(a)$ denotes the set of points in $\sigma_1 < \sigma < c$ at which $f(s)$ takes the value a, and $E_f(b)$ the set in $\sigma_1 < \sigma < c$ on which $f(s)$ takes the value b. Suppose that at most a finite number of points is taken from each of the sets $E_f(a)$ and $E_f(b)$ to obtain the sets $E'_f(a)$ and $E'_f(b)$. Let O' be any domain bounded connectedly with

respect to $\sigma > c$ which is at positive distance from $S_f + E'_f(a) + E'_f(b) + P_1$. If O' contains any points of $E_f(a) + E_f(b)$, extract from O' closed, non-overlapping circles interior to O' about each of these points; let O be the domain remaining. Then O is bounded connectedly with respect to $\sigma > c$ and is at positive distance from each point of $S_f + E_f(a) + E_f(b) + P_1$, so that $f(s)$ is bounded in O. Also, $f(s)$ is bounded in each of the extracted circles, which are closed, bounded sets in each of which $f(s)$ is holomorphic. Therefore $f(s)$ is bounded in O'. By definition, then, $f(s)$ is "M" in $\sigma > \sigma_1$ with respect to $S_f + E'_f(a) + E'_f(b)$. If $E_f(a) + E_f(b)$ contains at most a finite number of points, the above discussion applies with $E'_f(a)$ and $E'_f(b)$ both empty, so that $f(s)$ is "M" in $\sigma > \sigma_1$.

Lemma 1 of [2], referred to above, may be regarded as stating that *any function $f(s)$ given by a Dirichlet series having a finite abscissa of absolute convergence is "M" in $\sigma > \sigma_1$ with respect to $S_f + E_f(a) + E_f(b)$, where a and b are any two distinct complex numbers.*

S. Mandelbrojt ([4]; page 4) has given examples of functions bounded except for singularities, which are therefore also "M." Examples of "M" functions which are not bounded except for singularities are given in [2].

Consider two functions

$$f(s) = \sum a_n e^{-\lambda_n s}, \qquad \phi(s) = \sum b_n e^{-\lambda_n s}, \tag{2}$$

and the composite function

$$H(f, \phi) = H(s) = \sum a_n b_n e^{-\lambda_n s}. \tag{3}$$

Let S_ϕ denote the set of singular points of $\phi(s)$. Let c and c' be real numbers greater than α_f and α_ϕ respectively. Let A be a closed set in $\sigma < c$ containing S_f, and B a closed set in $\sigma < c'$ containing S_ϕ. Define

$$\begin{aligned}
\sigma(A) &= \max\Big[\sigma_1, \operatorname*{l.\,u.\,b.}_{\sigma+i\tau \in A,\ -\infty<\tau<\infty} \sigma\Big],\\
\sigma^\tau(A) &= \max\Big[\sigma_1, \operatorname*{l.\,u.\,b.}_{\sigma+i\tau \in A} \sigma\Big],\\
\sigma^\tau(B) &= \max\Big[\sigma_1, \operatorname*{l.\,u.\,b.}_{\sigma+i\tau \in B} \sigma\Big],\\
\sigma^\infty(A) &= \max\Big[\limsup_{\tau\to\infty} \sigma^\tau(A), \limsup_{\tau\to-\infty} \sigma^\tau(A)\Big],\\
\sigma^\infty(B) &= \max\Big[\limsup_{\tau\to\infty} \sigma^\tau(B), \limsup_{\tau\to-\infty} \sigma^\tau(B)\Big],\\
\sigma^\infty &= \max\left[\sigma_1 + \sigma^\infty(B),\ \sigma_2 + \sigma^\infty(A)\right].
\end{aligned} \tag{4}$$

Let $[A, B]$ denote the closure of the set of all points $\alpha + \beta$, where $\alpha \in A$, $\beta \in B$. It is, by definition, a closed set. Let A_τ denote the intersection of A

with the quarter-plane $t \geq \tau, \sigma > \sigma_1$ of the complex s-plane, where $s = \sigma + it$: $A_\tau = A \cdot (t \geq \tau, \sigma > \sigma_1)$. Define

$$[A, B]^\infty = \lim_{\tau \to \infty} [A_\tau, B], \tag{5}$$

the intersection of all sets $[A_\tau, B]$ as τ takes on all real values, which is itself a closed set (see [2]).

The theorem (cf. Theorem 2 of [2]) on which most of the theorems of the present paper are based may now be stated:

THEOREM 1. *Let A be any closed set in $\sigma < c$ which contains S_f, and let B be any closed set in $\sigma < c'$ which contains S_ϕ. If $f(s)$ is "M" in $\sigma > \sigma_1$ with respect to A, and if $\phi(s)$ is "M" in $\sigma > \sigma_2$ with respect to B, then $H(s)$ is holomorphic in that part of the complement in $\sigma > \sigma^\infty$ of $[A, B]^\infty \cdot (\sigma > \sigma^\infty)$ which is connected with $\sigma > c + c'$. Moreover, $H(s)$ is "M" in $\sigma > \sigma^\infty$ with respect to $[A, B]^\infty$.*

Proof. The theorem as stated above is precisely Theorem 2 of [2], with the exception of the last sentence, which is not included there. It remains, then, only to prove that under the given conditions $H(s)$ is "M" in $\sigma > \sigma^\infty$ with respect to $[A, B]^\infty$.

The sets $A(\delta)$ and $B(\delta)$ will be defined as in [2]: $A(\delta)$ is the complement in $\sigma \geq \sigma_1$ of the set $CA(\delta)$ defined as follows. Let $P(\delta)$ be a positive function of the positive variable δ, monotonely increasing to infinity as δ approaches zero. The set $CA(\delta)$ is the set of all curves of length less than $P(\delta)$, having one end-point in $\sigma > c$, which are central lines of channels in $\sigma > \sigma_1$ of width greater than 2δ which contain no points of A. $B(\delta)$ is defined in a similar manner, replacing σ_1 by σ_2, c by c', and A by B.

Let T denote the closed set $[A, B]^\infty$, which lies in the half-plane $\sigma \leq c + c'$. Define $T(\delta)$ as $A(\delta)$ was defined, replacing σ_1 by σ^∞, c by $c + c'$, and A by T. It is clear that every domain which is bounded connectedly with respect to $\sigma > c + c'$ and which is at positive distance from the set $T + P$, where P is the half-plane $\sigma \leq \sigma^\infty$, is interior to a domain $CT(\delta)$ for some positive δ. Also, every domain $CT(\delta)$ is itself bounded connectedly with respect to $\sigma > c + c'$, and is at positive distance from the set $T + P$. In order to show that $H(s)$ is "M" in $\sigma > \sigma^\infty$ with respect to T, it is therefore sufficient to show that $H(s)$ is bounded in each domain $CT(\delta)$.

Let δ' and δ'' be positive numbers such that $\delta' + \delta'' < \delta$, and write $\epsilon = [\delta - (\delta' + \delta'')]/5$. Let $M_f(\delta')$ be the least upper bound of $|f(s)|$ in $CA(\delta')$, and $M_\phi(\delta'')$ the least upper bound of $|\phi(s)|$ in $CB(\delta'')$. It can

easily be seen, by considering Mandelbrojt's integral expression for $H(s)$ (see [2], Theorem 1) that $H(s)$ is bounded in $CT(\delta)$. If $M_H(\delta)$ denotes the least upper bound of $|H(s)|$ in $CT(\delta)$, it can be shown that actually there exists a constant N independent of δ, δ', and δ'', such that

$$M_H(\delta) = M_H(5\epsilon + \delta' + \delta'') < (N/\epsilon) M_f(\delta') M_\phi(\delta''). \tag{6}$$

In order to show this, let z' be any point of $CT(\delta)$. There is then a curve L of length less than $P(\delta)$ joining z' to points in $\sigma > c + c'$, which is the central line of a channel in $\sigma > \sigma^\infty$ of width greater than 2δ which contains no point of $[A, B]^\infty$. It follows from Lemma 2 of [2] that if δ is sufficiently small, and if $\tau > \tau_0$ for τ_0 sufficiently large, then L is also the central line of a channel C' of width greater than $2[\delta - (\delta' + \delta'')]$, or 10ϵ, in $\sigma > \sigma^\infty$, which contains no point of $[A_\tau(\delta'), B(\delta'')]$, where $A_\tau(\delta') = A(\delta') \cdot (t \geq \tau)$. There is therefore a domain Δ in $\sigma > \sigma^\infty + 4\epsilon$ containing L and also points with arbitrarily large real part, whose boundary is at distance greater than 4ϵ from every point of $[A_\tau(\delta'), B(\delta'')]$, for $\tau > \tau_0$, and for δ sufficiently small. Fix δ', δ'', δ, and τ, so that this is true.

Pave the strip $\sigma_1 \leq \sigma \leq c$ of the s-plane ($s = \sigma + it$) with squares of side ϵ. Extract from this strip those closed squares, as well as those bordering on them, which contain points of $A_{\tau_0}(\delta')$. Extract also those bordering on the vertical line $\sigma = \sigma_1$. Denote by D_ϵ that part of the region in the half-plane $\sigma > \sigma_1$ remaining which is connected with the half-plane $\sigma > c$. Define the sequence $\{\alpha_m\}$ and the symbol $C_\epsilon(\alpha_m)$ as in Theorem 1 of [2]. It follows from the discussion in the proof of that theorem that

$$H(z') = \lim_{m \to \infty} (1/i\alpha_m) \int_{C_\epsilon(\alpha_m)} f(s)\phi(z' - s)\, ds.$$

Also, $C_\epsilon(\alpha_m)$ is interior to $CA(\delta')$, so that on $C_\epsilon(\alpha_m)$, $|f(s)| \leq M_f(\delta')$; and for s on $C_\epsilon(\alpha_m)$, $z' - s$ is in $CB(\delta'')$, so that $|\phi(z' - s)| \leq M_\phi(\delta'')$. It is pointed out in [2] that there is a constant N independent of ϵ such that the length of $C_\epsilon(\alpha_m)$ is less than $N\alpha_m/\epsilon$. It follows that

$$|H(z')| < (1/\alpha_m) M_f(\delta') M_\phi(\delta'') (N\alpha_m/\epsilon),$$

$$|H(z')| < (N/\epsilon) M_f(\delta') M_\phi(\delta'').$$

But z' may be any point of $CT(\delta)$, so that

$$M_H(\delta) = M_H(5\epsilon + \delta' + \delta'') < (N/\epsilon) M_f(\delta') M_\phi(\delta''),$$

which completes the proof of the theorem.

Mandelbrojt (see [6]) has introduced the concept "ultimate positive

abscissa of holomorphism of $f(s)$ with respect to σ_1." This real number is the larger of the two numbers σ_1 and the upper limit as τ becomes infinite of the largest real number σ such that $\sigma + i\tau \epsilon S_f$, where S_f denotes the set of singular points of $f(s)$. A closely allied terminology will be used here. The *ultimate abscissa of holomorphism of* $f(s)$, $\mathfrak{H}^\infty(f)$, is defined as follows:

$$\mathfrak{H}^\infty(f) = \max\left[\limsup_{\tau\to-\infty}\ (\max_{\sigma+i\tau\epsilon S_f}\sigma),\ \limsup_{\tau\to\infty}\ (\max_{\sigma+i\tau\epsilon S_f}\sigma)\right]. \tag{7}$$

It will be noted that with the previously used notation,

$$\sigma^\infty(S_f) = \max\left[\sigma_1, \mathfrak{H}^\infty(f)\right].$$

THEOREM 2. *If* $f(s) = \Sigma a_n e^{-\lambda_n s}$ *is "*M*" in* $\sigma > \sigma_0$ *for some real number* σ_0 *less than* $\mathfrak{H}_f - \limsup (\log n)/\lambda_n$, *then* $\mathfrak{H}^\infty(f) \geq \mathfrak{H}_f - \limsup (\log n)/\lambda_n$.

In particular, since $\mathfrak{H}^\infty(f) \leq \mathfrak{H}_f$, if $\limsup (\log n)/\lambda_n = 0$, and if $f(s)$ is "M" in $\sigma > \sigma_0$ for some real number σ_0 less than $\mathfrak{H}_f$, then $\mathfrak{H}^\infty(f) = \mathfrak{H}_f$; that is, $f(s)$ has singular points, the absolute values of whose imaginary parts are arbitrarily large, in every open half-plane containing the half-plane $\sigma \geq \mathfrak{H}_f$ in its interior.

Proof. Suppose that the theorem is not true, i.e., that $\mathfrak{H}^\infty(f) < \mathfrak{H}_f - \limsup (\log n)/\lambda_n$. Choose σ_1 so that $\sigma_0 < \sigma_1 < \mathfrak{H}_f - \limsup (\log n)/\lambda_n$, and $\sigma_1 > \mathfrak{H}^\infty(f)$. Then $f(s)$ is "M" in $\sigma > \sigma_1$, and $\sigma^\infty(S_f) = \max(\sigma_1, \mathfrak{H}^\infty(f)) = \sigma_1$. Put $A = S_f$ in Theorem 1; then A_τ, or $S_f \cdot (t \geq \tau, \sigma > \sigma_1)$, is empty for τ sufficiently large, and $[A, B]^\infty$ is empty, whatever the set B may be. In Theorem 1, put $\phi(s) = \Sigma e^{-\lambda_n s}$. Choose the set B as indicated in Section 5 of [2], so that $\phi(s)$ is "M" in the entire plane with respect to B. Since

$$\alpha_\phi = \limsup\ (\log n)/\lambda_n \tag{8}$$

(see [1] or [7]), such a set B can be constructed as follows: let $\{t_k\}$ be any sequence of non-negative numbers, strictly increasing to infinity, such that $\lim_{k\to\infty}(t_{k+1} - t_k) = 0$. Define $B = S_\phi + B_1$, where S_ϕ is the set of all singular points of $\phi(s)$, and B_1 is the set of points $s_k = \alpha_\phi + it_k$ and $s'_k = \alpha_\phi - it_k$ $(k = 1, 2, \cdots)$. Then $\phi(s)$ is "M" in $\sigma > \sigma_1$ with respect to B for $-\sigma_2$ arbitrarily large. The number σ_2 can be chosen so that $\sigma^\infty = \max[\sigma_1 + \sigma^\infty(B), \sigma_2 + \sigma^\infty(A)] = \sigma_1 + \sigma^\infty(B) = \sigma_1 + \alpha_\phi = \sigma_1 + \limsup (\log n)/\lambda_n$. Since $\sigma_1 < \mathfrak{H}_f - \limsup (\log n)/\lambda_n$, $\sigma^\infty < \mathfrak{H}_f$. But $[S_f, B]^\infty$ is empty, so that it follows from Theorem 1 that $H(s) = \Sigma a_n e^{-\lambda_n s} = f(s)$ is holomorphic in $\sigma > \sigma^\infty$. This evidently contradicts the above statement that $\mathfrak{H}_f > \sigma^\infty$, and completes the proof of the theorem.

THEOREM 3. *If* $f(s) = \Sigma a_n e^{-\lambda_n s}$ *is "M" in* $\sigma > \sigma_1$ *for some real number* $\sigma_1 < \mathcal{A}_f - \limsup (\log n)/\lambda_n$, *then* $\mathcal{H}_f \geq \mathcal{A}_f - \limsup (\log n)/\lambda_n$.

Proof. Suppose that the theorem is not true, i. e., that $\mathcal{H}_f < \mathcal{A}_f - \limsup (\log n)/\lambda_n$. Let d be any real number less than $\mathcal{A}_f \limsup (\log n)\lambda_n$ and greater than both σ_1 and $\mathcal{H}_f$. Since $\sigma_1 < d$, and since $f(s)$ has no singular points in $\sigma \geq d$, it follows from the definition of an "M" function that $f(s)$ is bounded by some positive number N in $\sigma \geq d$. Use of this fact with Cauchy's theorem in a classical manner yields the result that

$$\lim_{T\to\infty} (1/iT) \int_{d+0i}^{d+iT} f(s) e^{\lambda_n s}\, ds = \lim_{T\to\infty} (1/iT) \int_{c+0i}^{c+iT} f(s) e^{\lambda_n s}\, ds,$$

where c is a real number greater than $\mathcal{A}_f$, and the path of integration in each case is the straight line joining the end-points. But, as is well-known (e. g. [7], page 15),

$$a_n = \lim_{T\to\infty} (1/iT) \int_{c+0i}^{c+iT} f(s) e^{\lambda_n s}\, ds.$$

Also,

$$|a_n| = \left| \lim_{T\to\infty} (1/iT) \int_{d+0i}^{d+iT} f(s) e^{\lambda_n s}\, ds \right| \leq N e^{\lambda_n d}.$$

Hence

$$\Sigma |a_n e^{-\lambda_n z}| \leq \Sigma |a_n| \exp[-\lambda_n \mathcal{R}(z)] \leq N \Sigma \exp\{-\lambda_n [\mathcal{R}(z) - d]\}.$$

But by (8), $\Sigma e^{-\lambda_n s}$ converges if $\mathcal{R}(s) > \limsup (\log n)/\lambda_n$. Thus if $\mathcal{R}(z) > d + \limsup (\log n)/\lambda_n$, $\Sigma a_n e^{-\lambda_n z}$ converges absolutely. Therefore $\mathcal{A}_f \leq d + \limsup (\log n)/\lambda_n$. But $d < \mathcal{A}_f - \limsup (\log n)\lambda_n$, a contradiction which completes the proof of the theorem.

COROLLARY 1. *If* $f(s) = \Sigma a_n e^{-\lambda_n s}$ *is "M" in an open half-plane containing in its interior its axis of absolute convergence, and if* $\lim (\log n)/\lambda_n = 0$, *then* $\mathcal{H}^\infty(f) = \mathcal{H}_f = \mathcal{A}_f$.

COROLLARY 2. *If* $f(s)$ *is any function given by a Dirichlet series having a finite abscissa of absolute convergence, and if* $\sigma_1 < \mathcal{A}_f - 2 \limsup (\log n)/\lambda_n$ *(or if* $\sigma_1 < \mathcal{H}_f - \limsup (\log n)/\lambda_n$*) then either*

(i) $f(s)$ *has singular points, the absolute values of whose imaginary parts are arbitrarily large, in* $\sigma > \sigma_1$, *or*

(ii) $f(s)$ *takes infinitely often every value except at most one in* $\sigma_1 < \sigma < c$, *where* c *is any real number greater than* $\mathcal{A}_f$.

For if (ii) is not satisfied, then $f(s)$ is "M" in $\sigma > \sigma_1$, as is noted near the

beginning of this section. Then by Theorem 3, $\mathcal{H}_f \geq \mathcal{A}_f - \limsup (\log n)/\lambda_n$, so that $\sigma_1 < \mathcal{H}_f - \limsup (\log n)/\lambda_n$. By Theorem 2, $\mathcal{H}^\infty(f) \geq \mathcal{H}_f - \limsup (\log n)/\lambda_n > \sigma_1$, so that (i) is satisfied.

Two lemmas, which will be used in developing applications of the above theorems, follow.

LEMMA 1. *Let* $\{f_i(s)\}$ *be a finite set of functions,*

$$f_i(s) = \sum a_n^{(i)} \exp\{-\lambda_n^{(i)} s\}, \qquad (i = 1, 2, \cdots, N),$$

which satisfy the following conditions:

(i) $\max_{i=1,2,\ldots,N} \mathcal{A}_{f_i} = \mathcal{A} < \infty$;

(ii) $\lambda_m^{(i)} \neq \lambda_n^{(j)}$ *for all* m *and* n, *if* $i \neq j$;

(iii) $\lim (\log n)/\mu_n = 0$, *where the sequence* $\{\mu_n\}$ *is composed of the elements of all the sequences* $\{\lambda_n^{(i)}\}$, *arranged in increasing order;*

(iv) $F(s) = \sum_{i=1}^{N} f_i(s) = \sum b_n e^{-\mu_n s}$ *is "M" in* $\sigma > \sigma_0$ *for some real number* $\sigma_0 < \mathcal{A}$.

Then $\mathcal{H}_F = \mathcal{A}_F = \mathcal{A}$.

Proof. Under the hypotheses of the lemma (see [1] or [7]),

$$\mathcal{A}_F = \limsup \frac{\log |b_n|}{\mu_n} = \max_{i=1,2,\ldots,N} \limsup_{n\to\infty} \frac{\log |a_n^{(i)}|}{\lambda_n^{(i)}} = \mathcal{A}.$$

But by Corollary 1, $\mathcal{H}_F = \mathcal{A}_F$, which completes the proof.

LEMMA 2. *Suppose* $f(s) = \sum a_n e^{-\lambda_n s}$ *and* $\gamma(s) = \sum b_n e^{-\mu_n s}$ *are both "M" in* $\sigma > \sigma_1$ *for some real number* σ_1. *Let* $F(s)$ *denote their sum:* $F(s) = f(s) + \gamma(s)$. *Let* E *be the set of points (if any) in* $\sigma > \sigma_1$ *which are singular for both* $f(s)$ *and* $\gamma(s)$ *and at which* $F(s)$ *is holomorphic. If* E *is empty, or if the greatest lower bound of the distances between pairs of points of* E *is positive, then* $F(s)$ *is "M" in* $\sigma > \sigma_1$.

Proof. Let A denote the set S_f of singular points of $f(s)$ in $\sigma > \sigma_1$ and B the set S_γ of singular points of $\gamma(s)$ in $\sigma > \sigma_1$. Under the hypotheses, there exists a positive function $M_f(\delta)$ of the positive variable δ such that $|f(s)| < M_f(\delta)$ in $CA(\delta)$, which is defined near the beginning of this section. Similarly, there exists a function $M_\gamma(\delta)$ such that $|\gamma(s)| < M_\gamma(\delta)$ in $CB(\delta)$,

defined as is $CA(\delta)$, replacing A by B, and c, a number greater than α_F, by c_1, a number greater than α_γ. Let T be the set S_F of singular points of $F(s)$ in $\sigma > \sigma_1$, and define $CT(\delta)$ as $CA(\delta)$ was defined, replacing A by T, and c by $\max(c, c_1)$, which is certainly greater than α_F.

Let d be a positive number such that the distance between any two points of E is greater than d; let δ be any positive number less than $\frac{1}{4}d$, and let x be any point of $CT(\delta)$. There is then a curve L of length less than $P(\delta)$ (hence less than $P(\frac{1}{4}\delta)$) which joins x to $\sigma > \max(c, c_1)$ and is the central line of a channel C in $\sigma > \sigma_1$ of width greater than 2δ which contains no point of S_F. It can therefore contain no points of S_f or of S_γ except possibly points of E. There are at most a finite number n, depending only on δ, of points y_i $(i = 1, 2, \cdots, n)$ of E in C; for any channel whose central line is of length less than $P(\delta)$ can be covered by a finite number of circles of radius $\frac{1}{2}d$; each such circle contains at most one point of E. If y_i is any point of E, let K_i be a circle of radius $\frac{1}{4}\delta$ about y_i. Since $\delta < \frac{1}{4}d$, no two such circles intersect. If each K_i fails to intersect L, then L is the central line of a channel in $\sigma > \sigma_1$ of width greater than $\frac{1}{4}\delta$ which contains no point of A or B, so that x belongs to both $CA(\frac{1}{4}\delta)$ and $CB(\frac{1}{4}\delta)$, and $|f(x)| < M_f(\frac{1}{4}\delta)$, $|\gamma(x)| < M_\gamma(\frac{1}{4}\delta)$, therefore $|F(x)| < M_f(\frac{1}{4}\delta) + M_\gamma(\frac{1}{4}\delta)$.

If a circle K_i intersects L, K_i is at distance greater than $\frac{1}{2}\delta$ from the boundary of C.

If no K_i contains x, L may be replaced by a curve L', which is obtained by replacing every arc (if any) of L which is in a circle K_i by an arc of K_i between the two points of intersection. The length of L' is certainly less than $P(\delta) + 2n\pi\delta/4$, which is less than $P(k\delta)$ for some positive number k less than $1/8$, depending only on δ, since $P(\delta)$ increases monotonely to infinity as δ approaches zero. Moreover, L' joins x to $\sigma > \max(c, c_1)$, and is the central line of a channel in $\sigma > \sigma_1$ of width greater than $k\delta$ $(k < 1/8)$ which contains no point of A or B. Therefore $x \in CA(k\delta)$ and $x \in CB(k\delta)$. Hence $|F(x)| < M_f(k\delta) + M_\gamma(k\delta)$.

Finally, if a circle K_j contains x, there is only one such circle. By the discussion of the paragraph above, each point x' on the boundary of K_j lies in $CA(k\delta)$ and $CB(k\delta)$. Therefore $|F(x')| < M_f(k\delta) + M_\gamma(k\delta)$ for each such point. But $F(s)$ is holomorphic in K_j, and therefore takes its maximum value on the boundary, so that $|F(x)| < M_f(k\delta) + M_\gamma(k\delta)$.

This last inequality has been shown to hold for each point x in $CT(\delta)$, for each positive δ (k depending on δ). Thus $F(s)$ is bounded in each domain $CT(\delta)$; as has been noted before, this is sufficient in order that $F(s)$ be "M" in $\sigma > \sigma_1$. This completes the proof of the lemma.

The generalization for the sum of a finite number of functions is immediate. If each of the functions is "M," if E is the set of points singular for some or all of the functions on which the sum is holomorphic, and if the greatest lower bound of the distances between pairs of points of E is positive, then the sum also is "M."

2. Applications. In this section, theorems based on the theorems and lemmas of the previous section will be proved. First, it will have been noticed that the composite function $H(f, \phi)$ of Theorem 1 is defined only if the functions $f(s)$ and $\phi(s)$ are of the form: $f(s) = \Sigma a_n e^{-\lambda_n s}$, $\phi(s) = \Sigma b_n e^{-\lambda_n s}$, the two series having the same sequence of exponents. If, however, $f(s) = \Sigma a_n e^{-\lambda_n s}$ and $\phi(s) = \Sigma b_n e^{-\mu_n s}$, there is a minimum sequence $\{\nu_n\}$ which contains both sequences, $\{\lambda_n\}$ and $\{\mu_n\}$. The functions $f(s)$ and $\phi(s)$ can each be represented by a Dirichlet series whose sequence of exponents is $\{\nu_n\}$:

$$f(s) = \sum A_m e^{-\nu_m s}; \qquad A_m = a_n \quad \text{if} \quad \nu_m = \lambda_n;$$

$$A_m = 0 \quad \text{if} \quad \nu_m \neq \lambda_n \text{ for all } n;$$

$$\phi(s) = \sum B_m e^{-\nu_m s}; \qquad B_m = b_n \quad \text{if} \quad \nu_m = \mu_n;$$

$$B_m = 0 \quad \text{if} \quad \nu_m \neq \mu_n \text{ for all } n.$$

If the sequences $\{\lambda_n\}$ and $\{\mu_n\}$ have an infinite sequence $\{r_n\} = \{\lambda_n\} \cdot \{\mu_n\}$ in common ($\{r_n\}$ is the sequence of all non-negative numbers, arranged in increasing order, each of which appears in both $\{\lambda_n\}$ and $\{\mu_n\}$), if $r_i = \lambda_{n_i} = \mu_{m_i}$ $(i = 1, 2, \cdots)$, and if

$$f_1(s) = \sum a_{n_i} e^{-\lambda_{n_i} s}, \qquad \phi_1(s) = \sum b_{m_i} e^{-\mu_{m_i} s},$$

then

$$H(f, \phi) = H(f_1, \phi_1) = \sum A_m B_m e^{-\nu_m s} = \sum a_{n_i} b_{m_i} e^{-r_i s}.$$

If $\{\lambda_n\}$ and $\{\mu_n\}$ have no infinite sequence in common, the result obtained by the use of Theorem 1 is of course trivial, since then $H(f, \phi)$ is entire or constant. Use will be made in this section of Theorem 1, applied to functions whose sequences of exponents are not identical, but have an infinite subsequence in common. One such result follows, indicating that under certain conditions, two "M" functions represented by Dirichlet series can not each "cancel" the other's singularities so that their sum is holomorphic, unless they have essentially the same sequences of exponents.

THEOREM 4. *Suppose that* $F(s) = \Sigma A_n e^{-\lambda_n s}$ *and* $\Phi(s) = \Sigma B_n e^{-\mu_n s}$ *satisfy the following conditions:*

(i) $F(s)$ *is "M" in* $\sigma > \alpha_F - \eta$, $\Phi(s)$ *is "M" in* $\sigma > \alpha_\Phi - \eta$, *for some positive number* η;

(ii) $F(s) - \Phi(s)$ *is holomorphic and "M" in* $\sigma > \alpha_F - \eta$;

(iii) $\lim (\log |A_n|)/\lambda_n = \alpha_F$,

$\lim (\log |B_n|)/\mu_n = \alpha_\Phi$;

(iv) $\lim (\log n)/\lambda_n = \lim (\log n)/\mu_n = 0$;

Then the sequences $\{\lambda_n\}$ *and* $\{\mu_n\}$ *are identical, with the exception of at most a finite number of terms of each sequence.*

It is evident from Corollary 1 of Theorems 2 and 3 that under hypotheses (i) and (iv), $\mathcal{H}_\Phi = \alpha_\Phi$ and $\mathcal{H}_F = \alpha_F$. It follows from hypothesis (ii) that $\alpha_F = \alpha_\Phi$.

Incidentally, it will be observed that Lemma 2 assures the fulfillment of condition (ii), if (i) is satisfied, if $F(s) - \Phi(s)$ is holomorphic in $\sigma > \alpha_F - \eta$, and if further the greatest lower bound of the distances between pairs of points, singular for both $F(s)$ and $\Phi(s)$ and at which $F(s) - \Phi(s)$ is holomorphic, is positive.

Proof. It will first be shown that all of the terms of the sequence $\{\lambda_n\}$ except at most a finite number belong also to the sequence $\{\mu_n\}$. Write

$$f(s) = F(s + \alpha_F) = \Sigma A_n \exp [-\lambda_n (s + \alpha_f)] = \Sigma a_n e^{-\lambda_n s},$$

where

$$a_n = A_n \exp [-\lambda_n \alpha_F];$$

and

$$\phi(s) = \Phi(s + \alpha_F) = \Sigma B_n \exp [-\mu_n (s + \alpha_F)] = \Sigma b_n e^{-\mu_n s},$$

where

$$b_n = B_n \exp [-\lambda_n \alpha_F].$$

Then the hypotheses of Theorem 4 become:

(i) $f(s)$ is "M" in $\sigma > -\eta$; $\phi(s)$ is "M" in $\sigma > -\eta$; $\alpha_F = \alpha_\Phi = 0$;

(ii) $f(s) - \phi(s)$ is holomorphic and "M" in $\sigma > -\eta$;

(iii) $\lim (\log |a_n|)/\lambda_n = \lim (\log |A_n \exp[-\lambda_n \alpha_F|])/\lambda_n$
$= \alpha_F - \alpha_F = 0$, $\lim (\log |b_n|)/\mu_n = 0$;

(iv) $\lim (\log n)/\lambda_n = \lim (\log n)/\mu_n = 0$.

Suppose now that there exists an infinite subsequence $\{\lambda_{n_i}\}$ of $\{\lambda_n\}$ such that $\lambda_{n_i} \neq \mu_m$ for all positive integers i and m. Let $\{\lambda_{n_i}\}$ be the maximum such subsequence of $\{\lambda_n\}$. Let $\{\lambda_{m_j}\}$ be the complement in $\{\lambda_n\}$ of $\{\lambda_{n_i}\}$: each λ_{m_j} belongs both to $\{\lambda_n\}$ and to $\{\mu_m\}$. Let $\{\mu_{n_j}\}$ be the subsequence of $\{\mu_n\}$ which is identical with $\{\lambda_{m_j}\}$, and let $\{\mu_{m_k}\}$ denote its complement in $\{\mu_n\}$: $\lambda_{m_j} = \mu_{n_j}$ $(j = 1, 2, \cdots)$.

Write $\gamma(s) = f(s) - \phi(s)$. Then $\gamma(s)$ is holomorphic and "M" in $\sigma > -\eta$. In Theorem 1, replace $\phi(s)$ by $\gamma(s)$, A by S_f, and B by S_γ, which is empty. Here $\sigma_1 = \sigma_2 = -\eta$, $-\eta \leq \sigma^\infty(A) \leq 0$, $\sigma^\infty(B) = -\eta$. Then $\sigma^\infty = \max[\sigma_1 + \sigma^\infty(B), \sigma_2 + \sigma^\infty(A)] \leq -\eta$. The set $[A, B]^\infty$ is empty, so that by Theorem 1 $H(s) = H(f, \gamma)$ is holomorphic and "M" in $\sigma > -\eta$. The function $\gamma(s)$ is given by:

$$\gamma(s) = \sum a_n e^{-\lambda_n s} - \sum b_n e^{-\mu_n s} = \sum a_{n_i} e^{-\lambda_{n_i} s} + \sum (a_{m_j} - b_{n_j}) e^{-\lambda_{m_j} s} - \sum b_{m_k} e^{-\mu_{m_k} s}.$$

Then

$$H(s) = H(f, \gamma) = \sum a_{n_i}{}^2 e^{-\lambda_{n_i} s} + \sum a_{m_j}(a_{m_j} - b_{n_j}) e^{-\lambda_{m_j} s}.$$

Under the hypotheses, it is easily verified that neither series has an infinite abscissa of absolute convergence. The function $H(s)$ satisfies the conditions on $F(s)$ of Lemma 1, with

$$f_1(s) = \sum a_{n_i}{}^2 e^{-\lambda_{n_i} s}, \qquad f_2(s) = \sum a_{m_j}(a_{m_j} - b_{n_j}) e^{-\lambda_{m_j} s}, \qquad N = 2.$$

Also,

$$\alpha_{f_1} = \limsup (1/\lambda_{n_i}) \log |a_{n_i}{}^2| = 2 \limsup (-1/\lambda_{n_i}) \log |a_{n_i}| = 0,$$

by hypothesis (iii). Hence

$$\alpha = \max_{i=1,2} \alpha_{f_i} \geq 0,$$

so that $\mathfrak{M}_H \geq 0$, by Lemma 1. This contradicts the previous assertion that $H(s)$ is holomorphic in $\sigma > -\eta$, and completes the proof of the assertion that all of the terms of the sequence $\{\lambda_n\}$, except at most a finite number, belong also to the sequence $\{\mu_n\}$.

Upon interchanging $F(s)$ and $\Phi(s)$ in the discussion above, it becomes apparent that the two sequences $\{\lambda_n\}$ and $\{\mu_n\}$ are identical, with the exception of at most a finite number of terms of each sequence.

Hypothesis (iii) of Theorem 4 is certainly essential. For any series satisfying a condition of type (iv) and having a sufficiently small abscissa of absolute convergence might be added to $F(s)$, creating a new function such that all of the other hypotheses would be satisfied without the conclusion being valid.

As part of his thesis [3], Mandelbrojt found sufficient conditions in order that a Taylor series have at least two singular points on its circle of convergence. His method of approach there suggested the ideas developed in the next few paragraphs.

Throughout the rest of this section, two points will be said to belong to the same "class" (modulo $2\pi i$) if each may be obtained from the other by the addition of an integral multiple of $2\pi i$. Each class of points has a representative in any horizontal, half-open strip of width 2π: that point of the strip from which each point of the class may be obtained by the addition of a suitable integral multiple of $2\pi i$. In this way a many-one correspondence may be set up between points of any set in the plane, and points of a "representative" set in a given, horizontal, half-open strip of width 2π.

Consider the set of those singular points of a given function $f(s)$ which lie in the half-plane $\sigma > \sigma_1$. Let $R_f(\sigma > \sigma_1)$ denote the corresponding representative set (modulo $2\pi i$) in the strip $0 \leq t < 2\pi$. To each point of $R_f(\sigma > \sigma_1)$ corresponds one or more singular point of $f(s)$ with the same real part, whose imaginary parts differ from its imaginary part by integral multiples of 2π.

Let $R_f(\sigma = \sigma_0)$ denote the set representative of all points of S_f with real part σ_0. It consists of a point or points on the line segment $\sigma = \sigma_0$, $0 \leq t < 2\pi$.

This notation is useful in stating the following theorem, giving sufficient conditions in order that $R_f(\sigma = \alpha_F)$ contain at least two points. The following definition will also be required.

DEFINITION. The poles of a function in a region are said to have *finite maximum order* if there exists a positive integer K such that the order of every pole in the region is less than or equal to K. The smallest such integer is the maximum order, M, of the poles.

THEOREM 5. *Suppose that* $F(s) = \Sigma A_n e^{-\lambda_n s}$ *satisfies the following conditions:*

(i) $F(s)$ *is "*M*" in* $\sigma > \alpha_F - \eta$ *for some positive number* η;

(ii) $F(s)$ *has a singular point,* s_0, *on* $\sigma = \alpha_F$; *also, all singular points*

of $F(s)$ in $\sigma > \alpha_F - \eta$ lie on $\sigma = \alpha_F$, those at points $s_0 + 2\pi ki$ (k an integer) being poles of finite maximum order M;

(iii) *there exists a subsequence $\{\lambda_n^{(0)}\}$ of $\{\lambda_n\}$ such that $\{\lambda_n^{(0)}\} \neq \lambda_m + k$ for all m and n, and for all positive integers k less than or equal to M; and such that* $\limsup (\log |A_n^{(0)}|)/\lambda_n^{(0)} = \alpha_F$, *where $A_n^{(0)}$ is the coefficient of the* $\exp\{-\lambda_n^{(0)} s\}$ *term in the series $\Sigma A_n e^{-\lambda_n s}$;*

(iv) $\lim (\log n)/\lambda_n = 0$.

Then $R_F(\sigma = \alpha_F)$ contains at least two points.

Proof. Write

$$f(s) = F(s + s_0) = \Sigma A_n e^{-\lambda_n (s+s_0)} = \Sigma a_n e^{-\lambda_n s},$$

where $a_n = A_n e^{-\lambda_n s_0}$. Then $\alpha_F = 0$, and $f(s)$ satisfies the following conditions:

(i) $f(s)$ is "M" in $\sigma > -\eta$;

(ii) the point $s = 0$ is singular for $f(s)$; also, all singular points of $f(s)$ in $\sigma > -\eta$ lie on $\sigma = 0$, those at points $2\pi ki$ (k an integer) being poles of finite maximum order M;

(iii) $\limsup (\log |a_n^{(0)}|)/\lambda_n^{(0)}$
$= \limsup (1/\lambda_n^{(0)}) \log \{|A_n^{(0)}| \exp [-\lambda_n^{(0)} \alpha_F]\} = 0.$

It is desired to prove that $R_f(\sigma = 0)$ contains at least two points. Suppose that this is not true. Then the only singular points of $f(s)$ in $\sigma > -\eta$ are poles of finite maximum order M at certain points $2\pi ki$. Write

$$f_1(s) = f(s)(1 - e^{-s})^M = f(s) \sum_{j=0}^{M} (-1)^j C_{j,M} e^{-js}$$

$$= \sum_{n=0}^{\infty} a_n e^{-\lambda_n s} \sum_{j=0}^{M} (-1)^j C_{j,M} e^{-js} = \sum_{j=0}^{M} (-1)^j C_{j,M} \sum_{n=0}^{\infty} a_n e^{-(\lambda_n + j)s},$$

where $C_{j,M}$ is the binomial coefficient.

The function $f_1(s)$ may thus be expressed as the sum of a finite number of functions, each of which is evidently "M" in $\sigma > -\eta$, being the product of an "M" function and an entire function bounded in $\sigma > -\eta$. Moreover, $f_1(s)$ is holomorphic in $\sigma > -\eta$, since the zeros of $(1 - e^{-s})^M$ "cancel" the poles of $f(s)$. By Lemma 2 and the remark following, $f_1(s)$ is "M" in $\sigma > -\eta$.

Denote by $\{\lambda_n^{(j,k,\ldots,p)}\}$ the sequence of λ_n's (if any), each of which

belongs to the sequences $\{\lambda_n\}$, $\{\lambda_n + j\}$, $\{\lambda_n + k\}, \cdots, \{\lambda_n + p\}$, and to no others:

$$\{\lambda_n^{(j,k,\cdots,p)}\} = \{\lambda_n\} \cdot \{\lambda_n + j\} \cdot \{\lambda_n + k\} \cdots \{\lambda_n + p\};$$

$j, k, \cdots, p$ are positive integers less than or equal to M. For example, $\{\lambda_n^{(0)}\}$ denotes the sequence of positive numbers which belong to $\{\lambda_n\}$ but to no $\{\lambda_n + k\}$ $(k = 1, 2, \cdots, M)$; $\{\lambda_n^{(2)}\}$ is the sequence of λ_n's each of which belongs only to $\{\lambda_n\}$ and to $\{\lambda_n + 2\}$; $\{\lambda_n^{(1,2)}\}$ is the sequence of λ_n's, each of which belongs only to $\{\lambda_n\}$, $\{\lambda_n + 1\}$, and $\{\lambda_n + 2\}$. Designate by $a_n^{(j,k,\cdots,p)}$ the coefficient of $\exp\{-\lambda_n^{(j,k,\cdots,p)} s\}$.

The sum of series representing $f_1(s)$ may be rewritten as a Dirichlet series: $f_1(s) = \Sigma d_n e^{-\mu_n s}$. If $\{\mu_{n_i}\}$ is the sequence of all terms of $\{\mu_n\}$ which do not belong to $\{\lambda_n\}$, then $f_1(s)$ may be written:

$$\begin{aligned} f_1(s) = {} & \sum d_{n_i} e^{-\mu_{n_i} s} + \sum a_n^{(0)} e^{-\lambda_n^{(0)} s} + (1 - C_{1,M}) \sum a_n^{(1)} e^{-\lambda_n^{(0)} s} \\ & + (1 + C_{2,M}) \sum a_n^{(2)} e^{-\lambda_n^{(2)} s} + \cdots + [1 + (-1)^M C_{M,M}] \sum a_n^{(M)} e^{-\lambda_n^{(M)}} \\ & + (1 - C_{1,M} + C_{2,M}) \sum a_n^{(1,2)} e^{-\lambda_n^{(1,2)} s} + \cdots \end{aligned}$$

the sum of a finite number of series. Then

$$\begin{aligned} (9) \qquad H(s) = H(f, f_1) = {} & \sum [a_n^{(0)}]^2 e^{-\lambda_n^{(0)} s} \\ & + [1 - C_{1,M}] \sum [a_n^{(1)}]^2 e^{-\lambda_n^{(1)} s} + \cdots + \\ & + [1 + (-1)^M C_{1,M}] \sum [a_n^{(M)}]^2 e^{-\lambda_n^{(M)} s} \\ & + (1 - C_{1,M} + C_{2,M}) \sum [a_n^{(1,2)}]^2 e^{-\lambda_n^{(1,2)} s} + \cdots, \end{aligned}$$

the sum of a finite number of series. In Theorem 1, replace $\phi(s)$ by $f_1(s)$, A by S_f, and B by S_{f_1}, which is empty. Then $\sigma_1 = \sigma_2 = -\eta$, $\sigma^\infty(A) \leq 0$, $\sigma^\infty(B) = -\eta$, so that

$$\sigma^\infty = \max [\sigma_1 + \sigma^\infty(B), \sigma_2 + \sigma^\infty(A)] \leq -\eta.$$

By Theorem 1, $H(s) = H(f, f_1)$ is both "M" and holomorphic in $\sigma > -\eta$; $[A, B]^\infty$ is certainly empty since B is empty.

But $H(s)$ may be written as the sum of a finite number of functions as in (9). If

$$\gamma(s) = \sum [a_n^{(0)}]^2 e^{-\lambda_n^{(0)} s},$$

then, by hypotheses (iii) and (iv),

$$\begin{aligned} \alpha_\gamma &= \limsup (1/\lambda_n^{(0)}) \log |a_n^{(0)}|^2 \\ &= 2 \limsup (1/\lambda_n^{(0)}) \log |a_n^{(0)}| = 0. \end{aligned}$$

Thus at least one of the functions of which $H(s)$ is the sum has a zero abscissa of absolute convergence; by Lemma 1, $\mathfrak{A}_H = 0$. But this contradicts the previous assertion that $H(s)$ is holomorphic in $\sigma > -\eta$, and completes the proof of the theorem.

The purpose of supposing as hypothesis (iv) that $\lim (\log n)/\lambda_n = 0$ rather than that $\lim (\log n)/\lambda_n^{(0)} = 0$, was to assure that none of the functions of which $H(s)$ is a sum should have a positively infinite abscissa of absolute convergence.

The hypothesis $\limsup (\log |A_n^{(0)}|)/\lambda_n^{(0)} = \mathfrak{A}_F$ is certainly not impossible of fulfillment, in view of the fact that $\mathfrak{A}_F = \limsup (\log |A_n|)/\lambda_n$. It is essential, as the following example shows. Write $\phi(s) = \Sigma e^{-ns} = 1/(1-e^{-s})$. Define $g(s) = \Sigma c_n e^{-\mu_n s}$ so that $\lim (\log n)/\mu_n = 0$, $\mathfrak{A}_g < -\eta$, no μ_n is an integer, $\mu_n \neq \mu_m + 1$ for all n, m (for example, $\mu_n = n + 1/n$, $c_n = e^{-n}$). Write $F(s) = \phi(s) + g(s) = \Sigma A_n e^{-\lambda_n s}$; then $\mathfrak{A}_F = 0$. Hypotheses (i), (ii), and (iv) are satisfied. So also is the first part of (iii), the subsequence $\{\lambda_n^{(0)}\}$ being the sequence $\{\mu_n\}$. But

$$\limsup (1/\lambda_n^{(0)}) \log |A_n^{(0)}| = \limsup (\log |c_n|)/\mu_n \leq -\eta \neq 0;$$

and the conclusion is not valid. In fact, $R_F(\sigma = 0)$ contains only one point, $s = 0$. The theorem therefore fails if the hypothesis

$$\limsup (\log |A_n^{(0)}|)/\lambda_n^{(0)} = \mathfrak{A}_F$$

is not included.

Theorem 6, which follows, is an almost immediate consequence of Theorem 5.

THEOREM 6. *Suppose that* $F(s) = \Sigma A_n e^{-\lambda_n s}$ *is "M" in* $\sigma > \mathfrak{A}_F - \eta$ *for some positive* η, *and has as its only singular points in* $\sigma > \mathfrak{A}_F - \eta$ *poles of finite maximum order at certain points* $s_0 + 2\pi k i$ *for some point* s_0 *on* $\sigma = \mathfrak{A}_F$; *suppose further that* $\lim (\log n)/\lambda_n = 0$, *and that*

$$\lim (\log |a_n|)/\lambda_n = \mathfrak{A}_F.$$

Then there exists a positive integer n_0 *such that*

$$F(s) = \sum_{q=0}^{n_0} e^{-\lambda_q s} T_q(s);$$

where $T_q(s)$ *is a finite or infinite Taylor-Dirichlet series.*

Proof. All of the hypotheses of Theorem 5 are satisfied except (iii), and the conclusion does not hold. Also $\lim (\log | a_n |)/\lambda_n = \alpha_F$, so that $\limsup (\log | A_n^{(0)} |)/\lambda_n = \alpha_F$ is satisfied for any infinite subsequence $\{\lambda_n^{(0)}\}$. It follows that there exists no infinite subsequence $\{\lambda_n^{(0)}\}$ of $\{\lambda_n\}$ such that $\lambda_n^{(0)} \neq \lambda_m + k$ for all m and n, and for all positive integers k less than or equal to M. There then exists a positive integer n_0, such that for $n > n_0$, to each λ_n there corresponds a λ_{n_1} less than λ_n which differs from λ_n by a positive integer less than M. That is, to each λ_n $(n > n_0)$ there corresponds a λ_{n_1} and a positive integer R_{n_1} such that $0 < R_{n_1} \leq M$, and $\lambda_n = \lambda_{n_1} + R_{n_1}$. If $n_1 > n_0$, there is an n_2 such that $\lambda_{n_1} = \lambda_{n_2} + R_{n_2}$, where R_{n_2} is a positive integer, and $0 < R_{n_2} \leq M$. Evidently an n_k less than or equal to n_0 will be reached after repeating this a finite number of times, such that $\lambda_n = \lambda_{n_k} + (R_{n_1} + R_{n_2} + \cdots + R_{n_k})$. Thus to every n greater than n_0 there corresponds a positive integer $p(n) = n_k$, less than or equal to n_0, and a positive integer $R(n) = R_{n_1} + R_{n_2} + \cdots + R_{n_k}$ such that $\lambda_n = \lambda_{p(n)} + R(n)$. Hence to each positive integer q, less than or equal to n_0, there corresponds an empty, a finite, or an infinite sequence of λ_n's, each term of which is obtained from the preceding by adding a positive integer; the members of these sequences exhaust the sequence $\{\lambda_n\}$. That is, the sequence $\{\lambda_n\}$ is composed of a finite number n_0 of finite or infinite (or empty) sequences $\{\lambda_q + n_i^{(q)}\}$, where $\{n_i^{(q)}\}$ is, for each q less than or equal to n_0, a finite or infinite sequence of positive integers. Therefore $F(s)$ may be written:

$$F(s) = \sum_{q=0}^{n_0} e^{-\lambda_q s} \sum_{i=0}^{r_q} d_i^{(q)} e^{-n_i^{(q)} s},$$

where $0 \leq r_q \leq \infty$, and where the $d_i^{(q)}$ are complex numbers. On writing

$$T_q(s) = \sum_{i=0}^{r_q} d_i^{(q)} e^{-n_i^{(q)} s},$$

the form of the statement of the theorem is obtained.

3. Examples. In view of the fact that in most of the theorems of this paper the functions concerned are supposed "M" in some half-plane, it is pertinent to ask whether such functions exist. Obviously all periodic functions with an imaginary period have this property; in particular, those given by Taylor-Dirichlet series, which have period $2\pi i$. However, the class of functions which are "M" in open half-planes containing their half-planes of absolute convergence is far wider than that of periodic functions. For an example of a wide class of such functions, let $\phi(s)$ be given by a Taylor-

Dirichlet series $\Sigma b_n e^{-ns}$ having a finite abscissa of absolute convergence α_Φ. Let $f(s)$ be any function given by a Dirichlet series $\Sigma a_n e^{-\lambda_n s}$ such that $\alpha_f < \alpha_\Phi$. If $\alpha_f < \sigma_1 < \alpha_\Phi$, and if $\psi(s) = f(s) + \phi(s)$, then $\psi(s)$ is not only "M," but actually bounded except for singularities, in $\sigma > \sigma_1$. For certainly $\phi(s)$, being represented by a Taylor-Dirichlet series, is bounded except for singularities in $\sigma > \sigma_1$, and $f(s)$ is bounded in $\sigma > \sigma_1$, since $\sigma_1 > \alpha_f$. The function $\psi(s)$ thus furnishes the desired example. A much wider class of functions bounded except for singularities is given by Mandelbrojt ([4], page 4).

It is worth noting, in the example given above, that if $\psi(s) = \Sigma c_n e^{-\nu_n s}$, then $\lim (\log | c_n |)/\nu_n$ does not necessarily exist. Such an "M" function could not satisfy the conditions of Theorems 4 and 6. In fact, the sequence $(\log | c_n |)/\nu_n$ has at least two distinct limit points, if $\lim (\log n)/\lambda_n = 0$; for then $\alpha_f = \limsup (\log | a_n |)/\lambda_n$ is one limit point of the sequence $\{(\log | c_n |)/\nu_n\}$, and $\alpha_\Phi = \limsup (\log | b_n |)/n$ is another. To assure the non-triviality of Theorems 4 and 6, it will, therefore, be well to have an example of a non-periodic function $\Sigma c_n e^{-\nu_n s}$ which is "M" in an open half-plane containing in its interior its half-plane of absolute convergence, and for which $\lim (\log | c_n |)/\nu_n$ exists. An "M" function of the form $\Sigma e^{-\nu_n s}$ would satisfy this condition, for $\lim (\log 1)/\nu_n = 0$. In constructing such a function, let $\mu_1, \mu_2, \cdots, \mu_n$ $(M \geq 2)$ be a finite set of linearly independent positive numbers, so that $x_1\mu_1 + x_2\mu_2 + \cdots + x_M\mu_m = 0$, when $x_1, \cdots, x_M$ are integers, only if $x_1 = x_2 = \cdots = x_M = 0$. Then

$$x_1\mu_1 + x_2\mu_2 + \cdots + x_M\mu_m = y_1\mu_1 + \cdots + y_M\mu_m$$

for x_i and y_i $(i = 1, 2, \cdots, M)$ integers, only if $x_i = y_i$ $(i = 1, 2, \cdots, M)$. Write

$$f_j(s) = \sum e^{-\mu_j s} = \frac{1}{1 - e^{-\mu_j s}}, \qquad (j = 1, 2, \cdots, M).$$

Each $f_j(s)$ is periodic, with period $2\pi i/\mu_j$. Also, no $f_j(s)$ has a finite zero. Hence

$$F(s) = \prod_{j=1}^{M} f_j(s)$$

is singular at each finite point which is singular for some $f_j(s)$, and at no others. It follows that for every real number σ_1, $F(s)$ is bounded except for singularities in $\sigma > \sigma_1$, and $F(s)$ is "M" in the entire plane. But clearly $F(s) = \Sigma e^{-\nu_n s}$, where the set of non-negative numbers $\{\nu_n\}$ is the set of all

numbers of the form $x_1\mu_1 + x_2\mu_2 + \cdots + x_M\mu_M$, where each x_i $(i = 1, 2, \cdots, M)$ takes on all non-negative integral values. The function $F(s)$ is evidently not periodic, but is quasi-periodic in the sense of Bohl and Esclangon (cf. H. Bohr, "Zur Theorie der fast periodischen Funktionen," *Acta Math.*, vol. 46 (1925), p. 114, Theorem III).

THE RICE INSTITUTE.

BIBLIOGRAPHY

1. Bernstein, V., *Séries de Dirichlet*, Gauthier-Villars, Paris, 1933.
2. Brunk, H. D., "Theorems of Composition for Dirichlet Series," *Duke Mathematical Journal*, vol. 12 (1945), pp. 1-21.
3. Mandelbrojt, S., "Sur les séries de Taylor qui présentant des lacunes," *Annales de l'École Normale*, vol. 40 (1923), p. 428.
4. Mandelbrojt, S., "Contribution à la théorie du prolongement analytique des séries de Dirichlet," *Acta Math.*, vol. 55 (1930), pp. 1-32.
5. Mandelbrojt, S., "Sur la recherche des points singuliers d'une série de Dirichlet," *Bulletin de la Société Mathématique de France*, vol. 57 (1929), pp. 78-103.
6. Mandelbrojt, S., *Dirichlet Series*, The Rice Institute Pamphlet, 1944.
7. Valiron, G., "Théorie générale des séries de Dirichlet," *Mémorial des Sciences Mathématiques*, Gauthiers-Villars et C^{ie}, Paris, 1928.

EXTENSIONS OF PLANE CONTINUA MAPPINGS.*

By G. T. WHYBURN.

1. Introduction. Extension theorems for continuous transformations of the sort which extend the range of definition and retain continuity are well known—(see, for example, Chapter 1 of [1]). However, our knowledge of the extensibility of mappings so as to preserve given essential properties, such as monotoneity or interiority, in addition to continuity is much more limited. Some steps in this direction have been taken by the present author before (see Chapters 9 and 11 of [2]); and in this paper a further study of the problem is made. In particular, we shall deal largely with the situation in which a continuous transformation of a given sort acts on a locally connected continuum imbedded in a sphere (or a plane) and shall seek to extend the action of the mapping to the whole sphere without sacrificing its given vital properties. Indeed, in some cases it will be found possible to effect the extension so that the extended mapping is of a much stronger and more restricted type than was the original one.

All spaces used are assumed to be separable and metric. Frequently the space is of a much more restricted character and its nature is specified in the separate results.

If $A \subset X$, $B \subset Y$ and $f(A) = B$ and $g(X) = Y$ are continuous, g is said to be an extension of f to X provided

$$g(x) = f(x)$$

for all $x \in A$.

If $A \subset X$, $f(A) = B$ and $g(X) = Y$ are continuous and $B' = g(A)$, g is said to be an extension of f provided the transformations

$$f(A) = B \text{ and } g(A) = B'$$

are topologically equivalent, i. e., provided there exists a homeomorphism

$$h(B) = B'$$

such that

$$g(x) = hf(x), \qquad x \in A.$$

In other words, g is an extension of f provided we can imbed B topologically in Y so that

* Received April 25, 1945.

$$g(x) = f(x) \text{ for all } x \in A.$$

A continuous transformation $f(A) = B$ is said to be (i) *monotone*, if for each $y \in B$, $f^{-1}(y)$ is a continuum (i. e., compact and connected), (ii) *non-alternating* if for x, $y \in B$, $f^{-1}(y)$ separates no two points of $f^{-1}(x)$ in A, (iii) *interior* (or *open*) if the image of each open set in A is open in B, and (iv) *light* if for each $y \in B$, $f^{-1}(y)$ is totally disconnected.

A locally connected continuum is called (i) a *cactoid* (see [3]) if each of its true cyclic elements is a topological sphere, (ii) a *boundary curve* if each of its true cyclic elements is a simple closed curve, and (iii) a *dendrite* if it has no true cyclic elements.

For definitions of other terms used the reader is referred to the author's book *Analytic Topology* (see [2]).

2. Composite mappings on boundary curves. It is not true in general that non-alternating mappings applied successively yield a non-alternating composite transformation, although this is true of monotone and of interior transformations. However, for light non-alternating transformations acting on boundary curves this sort of "group property" was shown in [2], Chapter 9, to hold. In this section, we proceed to prove the validity of this conclusion without the assumption of lightness.

(2.1) LEMMA. *Let $f(A) = B$ be non-alternating where A (hence also B) is a boundary curve. Let E_b be any true cyclic element of B and let x, $y \in E_b$ be non cut points of B such that if $X = f^{-1}(x)$, $Y = f^{-1}(y)$, then $X + Y$ contains no cut point of A. Then $A - (X + Y)$ consists of just two components, namely, $f^{-1}(R)$ and $f^{-1}(S)$, where R and S are the two components of $B - (x + y)$.*

Proof. Let E_a be the true cyclic element of A such that $f(E_a) \supset E_b$. Then since neither X nor Y can separate A and neither contains a cut point of A, it results that X and Y are each continua and furthermore $X + Y \subset E_a$. Also it follows from this that $A - (X + Y)$ has just two components, say R' and S', each containing an open arc of E_a with one end point in X and the other in Y. Since $B - (x + y) = R + S$, it results that one of these, say R', maps onto R and the other, S', maps onto S.

(2.2) THEOREM. *If A is a boundary curve and the mappings $f_1(A) = B$ and $f_2(B) = C$ are non-alternating, so also is the mapping $f(A) = f_2 f_1(A) = C$.*

Proof. Since the non-alternating image of a boundary curve is a boundary curve, it results from our hypothesis that both B and C are boundary curves. Let $x, y \in C$, $f^{-1}(x) = X$, $f^{-1}(y) = Y$. We must show that X does not separate Y in A. Let $X' = f_2^{-1}(x)$, $Y' = f_2^{-1}(y)$ and let W be the least A-set in B containing X'.

Now if $W \cdot Y' = 0$, let Q be a component of $B - W$ intersecting Y' and let $p = F(Q)$. Then since p non ϵ Y', p cannot separate two points of Y' in B and hence $Y' \subset Q$. Thus since f_1 is non-alternating, $f_1^{-1}(Q)$ is a component of $A - f_1^{-1}(p)$ containing $f_1^{-1}(Y') = Y$ but containing no point of X. Thus X does not separate Y in this case.

If $W \cdot Y' \neq 0$, there exists a true cyclic element E_b of B intersecting both X' and Y'. Since f_2 is non-alternating, there exist disjoint arcs U and V of E_b containing $X' \cdot E_b$ and $Y' \cdot E_b$ respectively. Since the number of points of E_b which are either cut points of B or images under f_1 of cut points of A is countable, there exist points $a, b \in E_b$ which do not belong to this set and such that a and b separate U and V on E_b. Then $B - (a + b)$ consists of just two components R and S containing U and V respectively. Since X' can contain no cut point of B on $V \cdot E_b$ and Y' can contain no cut point of B on $U \cdot E_b$, we must have

$$X' \subset R, \qquad Y' \subset S.$$

Now by the lemma, $f^{-1}(R)$ and $f^{-1}(S)$ are connected and, since they contain $X = f_1^{-1}(X')$ and $Y = f^{-1}(Y')$, respectively, it results that X cannot separate Y in A.

3. Reduction to complementary domain closures. Since the general problem under study amounts to the question of how to define a mapping in the complementary domains of the continuum A on a plane or sphere which will be identical with the given mapping on A and will retain its essential properties on the whole sphere or plane, it is of first importance to know to what extent these properties automatically hold on the complementary domain boundaries of A when they hold on A. In this section this question is treated briefly for monotone and for non-alternating transformations. For interior transformations this question is treated in 7 along with the general extensibility problem.

(3.1) THEOREM. *Let $f(M) = A$ be non-alternating, where M is a plane locally connected continuum and A is a dendrite. Then if B is the boundary of any complementary domain of M, f is non-alternating on B.*

Proof. Let C be any true cyclic element of B, let $x, y \in f(C) = D$ and let z be an interior point of the arc xy of A which is of order 2 in A. Then since $A - z$ consists of two components one containing x and the other y and f is non-alternating, $M - f^{-1}(z)$ consists of two distinct components R and Q containing $X = f^{-1}(x)$ and $Y = f^{-1}(y)$ respectively.

Now if f alternated on C, for some $x, y \in f(C)$ there would exist $x_1, x_2 \in X \cdot C$ and $y_1, y_2 \in Y \cdot C$ such that x_1 and x_2 separate y_1 and y_2 on C. But since R contains an arc $x'_1x'_2$ spanning C where $x'_1 \in y_1x_1y_2$, $x'_2 \in y_1x_2y_2$ and this arc must intersect Q, we have the contradiction that $R \cdot Q \neq 0$. Hence the mapping $f(C) = D$ is non-alternating. Thus f is non-alternating on each true cyclic element of B.

Further, if $y \in f(B)$ and $a, b \in B \cdot f^{-1}(y)$, the set $E(a, b)$ of all points separating a and b in B is identical with the set of all points separating a and b in M. Thus since f is non-alternating on M, $E(a, b) \subset f^{-1}(y)$. Hence by [VIII, 2.4] p. 139 of [2], f is non-alternating on B.

(3.2) *Example.* As an interesting case where (3.1) applies, let us consider the mapping of the lemniscate L: $\rho^2 = \sin 2\theta$ onto the dendrite D composed of the intervals of the x and y-axes joining $(-1, 0)$ to $(1, 0)$ and $(0, -1)$ to $(0, 1)$ obtained as follows: map the points given by $\theta = 0$, $\theta = \pi/4$ and $\theta = 5\pi/4$ into the origin O. Then map the points given by $\theta = \pi/8$, $\theta = 3\pi/8$, $\theta = 9\pi/8$ and $\theta = 11\pi/8$ into the end points $p_1 = (1, 0)$, $p_2 = (0, 1)$, $p_3 = (-1, 0)$ and $p_4 = (0, -1)$ respectively of D. Finally map the remaining 8 segments of L by pairs topologically into the four open segments of $D - O - p_1 - p_2 - p_3 - p_4$ so as to get a continuous mapping f of L onto D. It is then readily observed that f is non-alternating on L and also on each loop of L as required by our theorem.

That (3.1) does not hold if no restriction is imposed on the image set A is seen by the following example. Let M consist of a circle C together with two perpendicular diameters x_1px_2 and y_1py_2 of C. Let f map M by identifying x_1 and x_2 and also y_1 and y_2 but otherwise being topological. Then clearly f is non-alternating on M but not on C, the boundary of the unbounded complementary domain of M.

However, in case f is monotone our same conclusion follows without any limitation on A, as will now be demonstrated.

(3.3) THEOREM. *Any monotone transformation on a plane (or spherical) locally connected continuum M is non-alternating on the boundary of each complementary domain of M.*

Proof. Let B be the boundary of any such complementary domain of M,

let C be any true cyclic element of B, let M' be the cyclic element of M containing C and let $f(C) = E \subset f(M) = A$. Now M' lies either in C plus its interior or in C plus its exterior, say in C plus its interior. Then if $x, y \in E$ and $X = M' \cdot f^{-1}(x)$, $Y = M' \cdot f^{-1}(y)$, X and Y are continua lying in C plus its interior. Let α be a minimum arc of C containing X. If $\alpha \cdot Y = 0$, then $Y \cdot C$ separates no two points of $X \cdot C$ in C. If $\alpha \cdot Y \neq 0$ then $Y \cdot C$ must lie wholly in α, since $X \cdot Y = 0$; in this case let β be the minimum subarc of α containing Y. Then $\beta \cdot X = 0$ since $X \cdot Y = 0$. Thus $C - \beta$ is a connected subset of $C - Y \cdot C$ containing $X \cdot C$; so that again $Y \cdot C$ separates no two points of $X \cdot C$ in C.

Thus f is non-alternating on each true cyclic element C of B. Further, for any $y \in A$ and any $a, b \in f^{-1}(y) \cdot B$, we must have $E(a, b) \subset f^{-1}(y)$ since f is non-alternating on M and any point separating a and b in B also separates them in M. Therefore, just as in the proof of (3.1), it follows that f is non-alternating on B.

4. Extensions from complementary domain closures. Once our definition of a given transformation on a plane or spherical continuum A is extended to the separate complementary domains of A so as to obtain the desired properties on the closures of these separate domains, the question arises as to what extent do these properties hold for the extended mapping on the whole sphere or plane. This section is devoted to the development of the needed results of this type for monotone and non-alternating transformations.

(4.1) THEOREM. *Any non-alternating transformation of a locally connected unicoherent continuum onto a dendrite is necessarily monotone.*

Proof. Let $f(A) = B$ be such a transformation. Let $y \in B$, $Y = f^{-1}(y)$ and suppose that Y is not connected. Then if $Y = Y_1 + Y_2$ is a separation of Y, there exists a closed subset X of $A - Y$ which separates $y_1 \in Y_1$ and $y_2 \in Y_2$ in A; and since A is unicoherent, we can suppose that X is a continuum. Now since $f(X)$ and y are disjoint subcontinua of the dendrite B, there exists a point z of B which separates y and $f(X)$ in B. But then $f^{-1}(z)$ separates y_1 and y_2 in A. For any arc y_1y_2 in A must intersect X and thus must intersect $f^{-1}(z)$, since $f^{-1}(z)$ separates Y and X in A.

Definition. A mapping $f(A) = B$ on a compact set A will be said to be monotone on a subset A_0 of A provided $A_0 \cdot f^{-1}(y)$ is either empty or connected for each $y \in B$.

(4.2) THEOREM. *Let M be a locally connected continuum on a sphere*

S and let $f(S) = D$ be a continuous mapping of S onto a dendrite D. Then if f is non-alternating on M and for each complementary domain R of M with boundary B, f is monotone on $\bar{R}$ and $f(\bar{R}) = f(B)$, then f is monotone on S.

Proof. By virtue of (4.1) we have only to show that f is non-alternating on S, because S is unicoherent. To this end, let $x, y \in D$ and let $X = f^{-1}(x)$, $Y = f^{-1}(y)$. Since f is non-alternating on M, there exists a component Q of $M - Y \cdot M$ containing $X \cdot M$. Let K be any component of X and let R be any complementary domain of M intersecting X. Then by hypothesis if $B = F(R)$, $X \cdot \bar{R}$ is a continuum and $f(\bar{R}) = f(B)$. Hence there exists a point $x_0 \in X \cdot B \subset X \cdot M$; and since $x_0 \in X \cdot \bar{R}$, we have $X \cdot \bar{R} \cdot Q \neq 0$. Since this holds for all complementary domains R of M which intersect X it follows that X lies in the component of $S - Y$ containing Q. Accordingly Y separates no two points of X on S; and hence f is non-alternating and thus monotone on S.

(4.3) THEOREM. *Let A be a locally connected continuum on a sphere S. Any monotone mapping $f(S) = K$ of S onto a dendrite K which is monotone on each complementary domain of A and on its closure is non-alternating on A.*

Proof. Let $f(A) = B$ and $x, y \in B$, $X = f^{-1}(x)$, $Y = f^{-1}(y)$. Then X is a continuum. Now for each complementary domain R of A with boundary F, since [1] f is non-alternating on F, there exists a component Q of $F - Y \cdot F$ containing $X \cdot F$. Hence if in each such case we replace $X \cdot R$ by Q in the set X we obtain a connected subset D of $A - A \cdot Y$ which contains $A \cdot X$. Accordingly it follows that the mapping $f(A) = B$ is non-alternating.

5. Monotone extensions of non-alternating transformations. It has been shown previously by the author (see p. 180 of [2]) that any non-alternating transformation f on a circle C can be extended monotonically to the interior I of C, i. e., so that the resulting mapping is monotone on $C + I$. Further, this extension can be effected in an extremely simple manner by decomposing $C + I$ into the convex closures $f^{-1}(y)_c$ of the sets $f^{-1}(y)$, for $y \in f(C)$, together with individual points of I not in any one of these sets $f^{-1}(y)_c$. Since other monotone extensions of f to I would exist, we shall refer

[1] This is a consequence of an unpublished result of the author's to the effect that a monotone mapping of the closure of a plane simply connected region R having property S onto a dendrite which is monotone on R is non-alternating on $F(R)$. See a forthcoming note, "Boundary alternation of monotone mappings," in the *Duke Mathematical Journal*.

to this particular one as the *monotone extension of f to I given by convexification of inverses.* This extension will serve as a starting point for most of the extensions to be developed in this paper. In general this extension involves an enlargement of the image space. However, in most cases which will concern us, the image space remains unchanged as indicated in the following theorem.

(5.1) THEOREM. *Let $f(C) = D$ be non-alternating where C is a circle and D is a dendrite and let ϕ be the monotone extension of f to the interior I of C given by convexification of inverses. Then $\phi(C + I) = D$.*

Proof. Let $E = I + C$. Then since ϕ is monotone, $\phi(E)$ is a hemicactoid with boundary curve D. Thus D is the base set of $\phi(E)$. Now if $\phi(E) - D \neq 0$, there would exist a point $y \in D$ and a component R of $\phi(E) - y$ containing no point of D. But then $\phi^{-1}(R)$ would be a component of $E - \phi^{-1}(y)$ not intersecting C. Clearly this is impossible since $\phi^{-1}(y)$ is convex.

We have the following interesting corollary of a geometric nature:

(5.11) COROLLARY. *If $f(C) = D$ is any non-alternating mapping of the circle C onto a dendrite D, any point within C is contained uniquely in the convex closure of $f^{-1}(y)$ for some $y \in B$.*

(5.2) LEMMA. *If A is a locally connected continuum which is the boundary of a plane (or spherical) bounded region R and $f(A) = D$ is any non-alternating mapping of A onto a dendrite D, there exists a monotone extension ϕ of f to $R + A$ such that $\phi(R + A) = D$ and which is monotone also on R.*

Proof. By [VIII, (9.8)], p. 161, of [2] there exists a non-alternating light mapping g of the circle C plus its interior I onto $A + R$, such that $g(C) = A$, $g(I) = R$ and the mapping $g(C) = A$ is non-alternating and the mapping $g(I) = R$ is topological. By (2.2) the mapping $fg(C) = D$ is non-alternating. Let h be the monotone extension of the mapping fg to I given by convexification of inverses. Since D is a dendrite, we have $h(C + I) = D$ by (5.1).

Now for each $x \in R + A$, let us define

$$\phi(x) = hg^{-1}(x).$$

Then if $x \in A$, we have

$$\phi(x) = fgg^{-1}(x) = f(x),$$

so that ϕ is an extension of f. Further for $y \in D$, $\phi^{-1}(y) = gh^{-1}(y)$; and since

$h^{-1}(y)$ is a continuum, it results that ϕ is monotone. Clearly $\phi(R+A) = h(C+I) = D$; and ϕ is monotone on R, since $I \cdot h^{-1}(y)$ is connected for each $y \in D$ and the mapping $g(I) = R$ is topological.

Since (5.2) effects the extension of f to the region R by reducing it back to the case where the boundary is a circle and then uses the "convexification of inverses" method, we shall speak of the extension ϕ obtained in (5.1) also as *the monotone extension given by convexification of inverses.*

(5.3) THEOREM. *Any non-alternating mapping $f(A) = D$ of a locally connected continuum A on a sphere S onto a dendrite D is monotonically extensible to S—and indeed so that the extended mapping is monotone on each complementary domain and on its closure and each such domain closure has the same image as the boundary of the domain.*

Proof. Let R be any complementary domain of A on S and let B be its boundary. By (3.1), f is non-alternating on B; and since $f(B)$ is a subdendrite of D, by (5.2) there exists a monotone extension ϕ of f to R so that $\phi(\bar{R}) = f(B)$. Let ϕ be defined on S as the mapping obtained by extending f to each such R in this way. Then clearly ϕ is monotone on each R and on the closure of each such R and $\phi(\bar{R}) = \phi(B)$. Further, since f is continuous, since $\phi(x) \equiv f(x)$ for $x \in A$ and since the complementary domains of A on S form a null sequence at most, it results that the mapping $\phi(S) = D$ is continuous. Finally, ϕ is non-alternating on A since it is identical with f on A; and therefore, by (4.2), ϕ is monotone on S. This completes the proof.

Definition. Since other monotone extensions of f to S would exist, we shall refer to the particular one set up in the preceding theorem and its proof as the *monotone extension given by the convexification of inverses method.*

Results (4.3) and (5.3) yield at once

(5.4) THEOREM. *In order that the continuous mapping $f(A) = D$ of a locally connected continuum A on a sphere S onto a dendrite D be non-alternating it is necessary and sufficient that f be monotonically extensible to S so that the extended mapping ϕ is monotone on each complementary domain of A and on its closure and so that $\phi(S) = D$.*

Thus if A is a locally connected continuum on a sphere S, the class $[f]$ of all non-alternating mappings of A into dendrites D is equivalent to the class $[\phi]$ of all monotone mappings on S which map S into dendrites and which are monotone on the complementary domains of A and on their closures. For each such $f \in [f]$ is associated uniquely with a $\phi \in [\phi]$ such that $f(x) \equiv \phi(x)$ for $x \in A$.

Remark. The reader may find it of interest to follow through the steps of the extension as it would be effected by the procedure described in (5.2) and (5.3) in the case of the example of the lemniscate mapping into a cross as defined in (3.2).

6. Monotone extensions of monotone transformations. Since any monotone transformation is non-alternating, it results from the theorems in **5** that any monotone transformation of a locally connected continuum on a sphere onto a dendrite is monotonically extensible to the whole sphere and, indeed, so that the extended mapping satisfies the extra conditions in (5.3). However, for monotone transformations the problem can be handled in a much simpler way.

For let A be any compact subset of a separable metric space M. Then *any monotone mapping* $f(A) = B$ *on* A *can be extended monotonically to* M. To do this it is only necessary to decompose M into the sets $f^{-1}(y)$, $y \in B$, together with the individual points of $M - A$; let M' be the hyperspace of this decomposition and let $\phi(M) = M'$ be the associated transformation. Then ϕ is monotone and, since ϕf^{-1} is a hor omorphism h mapping B onto a subset B' of M' so that $hf(x) \equiv \phi f^{-1} f(x) = \phi$) for $x \in A$, ϕ is an extension of f to M.

Since any monotone image of a ere is a cactoid, we have

(6.1) THEOREM. *Any monotone image of a compact set on a sphere is topologically contained in a cactoid.*

For locally connected continua A on a sphere S it is interesting to note that the method of convexification of inverses affords a second—and in general different—method of effecting monotone extensions of arbitrary monotone transformations on A, whether the image of A is a dendrite or not. For if f is monotone on A it is non-alternating on the boundary of each complementary domain of A by (3.3). Accordingly the method of convexification of inverses affords a continuous extension ϕ of f to S as described in the proof of (5.3), except that it is no longer assured that each complementary domain closure will have the same image as the boundary of the domain. However, for any $y \in \phi(S)$, $\phi^{-1}(y)$ either intersects A or else lies entirely in a single complementary domain of A; and since f is monotone and ϕ is monotone on each complementary domain closure, it follows in either case that ϕ is monotone.

It may be noted that when the former (more general) extension method is used, A would be an inverse set under the extended mapping, whereas when the method of convexification of inverses is used this is not usually the case. Both methods in general involve enlargement of the image space. It would be of interest to compare the two methods of extension more fully.

Although we have been able, in the case of monotone mappings, to dispense with the restriction that the image set be a dendrite, this cannot be done in the case of non-alternating transformations in general. Further, no significant assertion can be made concerning the structure of the non-alternating image of a locally connected continuum on a plane or sphere. These facts are exhibited by

(6.2) EXAMPLE. *There exists a non-alternating light mapping of a 2-cell, or of a regular curve A on a sphere S, onto a set containing the Q_ω which is topological on the boundary of each complementary domain of A and yet which cannot be extended monotonically to S.*

For let A be either a hemisphere on S or the sum of two Sierpinski regular curves [4] on S having a common side I_0 but otherwise being disjoint. In the former case let I denote a non-degenerate arc lying entirely within A and in the latter let I be a subinterval of I_0 containing neither end point of I_0. In either case let $f(I) = Q_\omega$ be any continuous light mapping. Decompose A into the sets $f^{-1}(y)$, $y \in Q_\omega$, together with the individual points of $A - I$ and let B be the hyperspace and $g(A) = B$ the associated transformation. Then B contains $gf^{-1}(Q_\omega)$ which is homeomorphic with Q_ω. Further, no set $g^{-1}(z)$, $z \in B$, separates A so that g is non-alternating.

Clearly in either case g is topological on the boundary of each complementary domain of A since no such boundary intersects I in more than one point. In neither case can there exist a monotone extension of g to S. For let $y \in g(A)$ such that $g^{-1}(y)$ is non-degenerate. Then $g^{-1}(y)$ is a totally disconnected subset of I and there exists no continuum in $S - A + g^{-1}(y)$ containing two distinct points of $g^{-1}(y)$.

7. Monotone extensions of interior transformations. In view of results already obtained the question naturally arises as to whether the monotone extension we have obtained for a given non-alternating mapping of a locally connected continuum on a sphere will be interior provided the original mapping is interior. This section is devoted to developing conditions under which this will be true.

(7.1) THEOREM. *Let $f(A) = B$ be interior and non-alternating where A is a plane (or spherical) locally connected continuum. Then for any cut point p of B, $f^{-1}(p)$ intersects the boundary F of any complementary domain D of A in at most two points which are accessible from a given component Q of $A - f^{-1}(p)$.*

Proof. For if three points x_1, x_2, x_3 of $F \cdot f^{-1}(p)$ were accessible from Q, $D + Q + x_1 + x_2 + x_3$ would contain a θ—curve $\theta = ax_1b + ax_2b + ax_3b$, where $a \in D$, $b \in Q$. But there exists a second component R of $A - f^{-1}(p)$; and by the non-alternating property, $R = f^{-1}f(R)$. Hence by interiority, $F(R) = f^{-1}(p)$. Thus $F(R) \supset x_1 + x_2 + x_3$ and $R \cdot \theta = 0$; and clearly this is impossible.

By an exactly similar argument we can prove

(7.2) THEOREM. *Let $f(A) = B$ be interior and non-alternating where A is a plane (or spherical) locally connected continuum. Then for any point p of B which cuts B into at least three components, at most two points of $f^{-1}(p)$ are accessible from each of two components of $A - f^{-1}(p)$.*

Since if a region has property S all its boundary points are accessible, we have the following corollary [assuming the same situation as in (7.1) or (7.2)].

(7.21) COROLLARY. *If for each $p \in B$, each component of $A - f^{-1}(p)$ has property S, then*

(a) *for any cut point p of B, $f^{-1}(p)$ intersects the boundary of each complementary domain of A in at most two points, and*

(b) *for any point p which cuts B into at least three components, $f^{-1}(p)$ consists of at most two points.*

Now if f is light, the condition in (7.21) is automatically satisfied. Hence we have

(7.22) COROLLARY. *If $f(A) = B$ is interior, non-alternating and light, where A is a locally connected continuum on a plane or a sphere, conclusions* (a) *and* (b) *or* (7.21) *hold.*

As a standing hypothesis for statements (1)-(11) which follow, let $f(A) = B$ be interior, light and non-alternating where A is a locally connected continuum on a sphere S and B is a dendrite.

(1) For any end point y of B, $f^{-1}(y)$ is a single point.

(2) For any branch point y of B, $f^{-1}(y)$ consists of at most two points.

(3) For any $y \in B$, $f^{-1}(y)$ intersects the boundary of each complementary domain of A in at most two points.

Proof. By (1) and (2) we need only consider the case where y cuts B into exactly two components. Hence $f^{-1}(y)$ cuts A into just two components

R and Q and every point of $f^{-1}(y)$ is accessible from both R and Q. Accordingly at most two points of $f^{-1}(y)$ could lie on the boundary of any given complementary domain of A.

(4) If C is a simple closed curve in the boundary of any complementary domain R of A, then $f(C)$ is a simple arc ab and $C \cdot f^{-1}(a)$ and $C \cdot f^{-1}(b)$ are single points, whereas $C \cdot f^{-1}(x)$, $x \in ab - a - b$, is a pair of points. Thus f is interior on C.

(5) If A is cyclic, f maps the boundary of each complementary domain of A interiorly and $(2-1)$ onto an arc of B.

(6) If no pair of points disconnects A into more than two components, B is necessarily a simple arc.

Proof. If B had a branch point y, by (2) $f^{-1}(y)$ would be a pair of points in A cutting A into at least three components.

Let $\phi(S) = B$ be the monotone extension of f to S, obtained by extending f monotonically in the complementary domain boundaries of A by *the convexification of inverse sets* process.

(7) For each end point y of B, $\phi^{-1}(y) = f^{-1}(y) =$ a single point.

(8) For each point y of B of order 2, $\phi^{-1}(y)$ is a simple closed curve.

Proof. Let R and Q be the two components of $A - f^{-1}(y)$. Since ϕ is monotone, $\phi^{-1}(y)$ is a continuum separating R and Q on S. Further, since the part of $\phi^{-1}(y)$ in any complementary domain of A is an open simple arc and $f^{-1}(y)$ is totally disconnected, $\phi^{-1}(y)$ is locally connected. Also since any complementary domain of A containing an arc $ab - (a+b)$ of $\phi^{-1}(y)$ is divided into two regions which are separated by $\phi^{-1}(y)$, it follows that $\phi^{-1}(y)$ separates R and Q irreducibly on S. Accordingly $\phi^{-1}(y)$ is a simple closed curve.

(9) If the image of each complementary domain boundary of A is a free arc in B, then ϕ is interior.

Proof. For if $x \in A$, and U is any neighborhood of x, $f(U \cdot A)$ has $f(x)$ as an interior point relative to B. Hence also $\phi(U)$ has $\phi(x) = f(x)$ as an interior point relative to B, since $\phi(U) \supset f(U \cdot A)$.

Also, if $x \in R$, a complementary domain of A, then since $R \cdot \phi^{-1}\phi(x)$ is an open arc dividing R into two regions R_1 and R_2 mapping under ϕ onto opposite sides of $\phi(x)$ in the arc $f[F(R)]$ and each having x on its boundary, it results that if U is any neighborhood of x, $\phi(U)$ has $\phi(x)$ as an interior point relative to B.

(10) If B is a simple arc, ϕ is interior.

(11) If the image of each complementary domain boundary of A is an arc, then f is interior on each complementary domain boundary.

Proof. It is non-alternating there and hence is interior, since any non-alternating light transformation onto an arc is interior.

(7.3) THEOREM. *Any light interior and non-alternating transformation $f(A) = ab$ of a locally connected continuum A on a sphere S onto a simple arc ab can be extended to a monotone interior mapping $\phi(S) = ab$. Further, $\phi^{-1}(a)$ and $\phi^{-1}(b)$ are single points while $\phi^{-1}(y)$ is a simple closed curve for any $y \in ab - a - b$.*

Now since any 1-dimensional locally connected continuum A which is identical with its cyclic chain $C(a, b)$, $a, b \in A$, can be mapped [5] onto an interval by a light non-alternating interior transformation, (7.3) yields the following result.

(7.31) EXISTENCE THEOREM. *If A is any 1-dimensional locally connected continuum on the sphere S which is identical with the cyclic chain $C(a, b)$ in A, there exists a monotone interior mapping ϕ of the whole sphere onto the interval $[0, 1]$ so that ϕ is light, interior and non-alternating on A and $\phi^{-1}(0) = a$, $\phi^{-1}(1) = b$ and for $0 < y < b$, $\phi^{-1}(y)$ is a simple closed curve separating a and b.*

The assumption of lightness is essential for the validity of (7.3) as is readily seen by a simple example. The extra condition needed in the absence of any lightness assumption is exhibited in the following theorem concerning monotone interior extensibility of the general non-alternating interior mappings in the situation under consideration.

(7.4) THEOREM. *An interior non-alternating mapping $f(A) = ab$ of a locally connected continuum A on a sphere S onto a non-degenerate simple arc ab can be extended to a monotone interior mapping $g(S) = ab$ of the whole sphere onto ab if and only if each component of $S - X$ intersects A, where X is any component of a set $f^{-1}(y)$, $y \in ab$.*

Proof. To establish the sufficiency of the condition, let $f(A) = B$ be interior and non-alternating and let

$$f = f_2 f_1, \quad f_1(A) = A' \text{ monotone}, \quad f_2(A') = ab \text{ light},$$

be the monotone—light factorization of f. Let

$$g_1(S) = K, \qquad K \supset A'$$

be the monotone extension of f_1 to S given by decomposing S into the sets $f^{-1}(y)$, $y \in A'$, and single points of $S - A$ and considering A' as imbedded in the hyperspace K of this decomposition (see **6**). Then K is a cactoid. Further, since $f^{-1}(a)$ and $f^{-1}(b)$ are connected, $a' = g_1 f^{-1}(a)$ and $b' = g_1 f^{-1}(b)$ are single points; and it results at once from our hypothesis that K is identical with the cyclic chain $C(a', b')$ in K.

For suppose there is a component R of $K - C(a', b')$. Then if $p = F(R)$, $g_1^{-1}(p)$ is a component X of $f^{-1}fg_1^{-1}(p)$; and since g_1 is monotone, $g_1^{-1}(R)$ contains a point x of A. But this means that $g_1^{-1}(R)$ contains either $f^{-1}(a)$ or $f^{-1}(b)$, say $f^{-1}(a)$, because every component of $A - f^{-1}fg_1^{-1}(p)$ contains either $f^{-1}(a)$ or $f^{-1}(b)$. Clearly this is impossible since then R would contain $g_1 f^{-1}(a) = a'$ which is absurd.

Now consider the mapping $f_2(A') = ab$. This mapping is non-alternating, interior and light. Further, if C_i is an arbitrary true cyclic element of $K = C(a', b')$, f_2 is non-alternating, light and interior on $A' \cdot C_i$.

Thus since the image $f_2(A' \cdot C_i)$ is a subarc $x_i y_i$ of ab, the mapping f admits of a monotone interior extension to all of C_i as given in (7.3). Let this extension be effected in each true cyclic element C_i of K and let $g_2(K) = ab$ be the resulting extension of f_2 to K. Since $E(a', b') \subset A'$ and since the arcs $x_i y_i$ can overlap only at their endpoints, clearly g_2 is single valued, continuous and monotone. Further, since $E(a', b')$ maps topologically [2] under f_2, it results at once that it also maps topologically under g_2 and thus that g_2 is interior, since g_2 is interior on each C_i.

Finally, let the decomposition of S into the sets $[g_1^{-1}g_2^{-1}(y)]$, $y \in ab$, generate the mapping $g(S) = ab$. Then g is monotone, since both g_1 and g_2 are monotone. Also g is identical with f on A, because $f = f_2 f_1$ and g_1 and g_2 are extensions of f_1 and f_2 respectively. Finally, g is interior. For since $g(A) = ab = g(S)$, g is interior at any point of A, and if x is in a component R of $S - A$, g_1 is topological and g_2 is interior on R so that g is interior at x, since $g = g_2 g_1$. This completes the sufficiency proof.

To prove the necessity, suppose that the mapping $f(A) = ab$ admits a monotone interior extension $g(S) = ab$ to the whole sphere S but that for some component X of $f^{-1}(y)$ for some $y \in ab$, there is a component R of $S - X$ with $S \cdot A = 0$. Since g is interior, $g(\bar{R}) \neq y$. Let $z \in g(\bar{R}) - y$. Then $g^{-1}(z) \cdot R \neq 0$ since $g(X) = y$; and $g^{-1}(z) \cdot (A - X) \neq 0$ since $g(A) \supset z$. But this

[2] This statement is easily verified. Also it is a consequence of a much more general result concerning non-alternating interior mappings. See the author's paper, "On uniqueness of the inverse of a transformation," *Duke Mathematical Journal*, vol. 12 (1945), pp. 317-323.

is impossible since $g^{-1}(z)$ is connected and does not intersect X. This completes the proof.

It should be noted that in case the given mapping f is light, the condition in (7.4) is automatically satisfied. Hence (7.4) includes (7.3) as a special case.

The following examples shed light on questions which arise in connection with extensibility of interior transformations on various curves lying on a sphere or on a plane.

(7.51) EXAMPLE. *There exists a light interior mapping of a θ-curve onto an interval which is not interior on the boundary of one of the complementary domains to the curve.*

For let A be a circle C together with the horizontal diameter ao_2 of C. Let the upper semicircle of C be lettered $ao_1b_1o_2$ and the lower one $ao_3b_2o_2$ by choosing points in the indicated orders. Let $a'o'b'$ be an interval. Then map A into $a'o'b'$ by sending a into a', o_1, o_2 and o_3 into o', b_1 and b_2 into b' and sending the residual open intervals of A topologically into either $a'o' - (a' + o')$ or $o'b' - (o' + b')$ so as to obtain a continuous mapping f of A into $a'o'b'$. Then f is light and interior but is not interior on C.

(7.52) EXAMPLE. *There exists a light interior non-alternating mapping of a θ-curve into a dendrite (indeed a T-curve) T which admits no monotone interior extension to the whole sphere so that the image is still T.*

For let A consist of a circle C together with the horizontal diameter o_1yo_2 where y is the center of C. Let xo_1z and xo_2z be the right and left semicircles respectively of C and let R be the upper region into which o_1yo_2 divides the interior of C. Let f map A into the part T of A to the right of and on the line xyz by reflection of the part of C to the left in this line.

Then f is interior, light and non-alternating but admits no extension ϕ of the sort described above. For if so then since R is connected and has x and y on its boundary, its image under ϕ must contain o_1. Hence there is a point u in R with $\phi(u) = o_1$. By interiority of ϕ there is a point v in R with $\phi(v) = v' \in o_1z - o_1$. But $\phi^{-1}(v')$ then intersects both R and o_1zo_2 but does not intersect the boundary $o_1xo_2 + o_1yo_2$ of R, contrary to the fact that it must be connected.

8. Conclusion. It may be remarked that nearly all our extensions in this paper have been effected by the "convexification of inverses" method and

in only a few cases is the image space augmented. Other means of securing extensions could surely be applied and would doubtless yield interesting and significant conclusions even for the types of transformations considered here. A study of the general extension question from this viewpoint is under way and will probably be dealt with in a later paper.

UNIVERSITY OF VIRGINIA.

BIBLIOGRAPHY.

[1]. Alexandroff-Hopf, *Topologie*, J. Springer, Berlin, 1935.

[2]. G. T. Whyburn, *Analytic Topology*, American Mathematical Society Colloquium Publications, New York, 1942.

[3]. R. L. Moore, "Concerning upper semi-continuous collections," *Monatshefte für Mathematik und Physik*, vol. 36 (1929), pp. 81-88.

[4]. W. Sierpinski, *Prace Matematyczno Fizyczne*, vol. 27 (1915); or Knaster and Kuratowski, *Bulletin of the American Mathematical Society*, vol. 33 (1927), pp. 106-107.

[5]. G. T. Whyburn, "The existence of certain transformations," *Duke Mathematical Journal*, vol. 5 (1938), pp. 647-655.

A THEORY OF CRITICAL POINTS.*

By MAGNUS R. HESTENES.

1. Introduction. Let $f(x) = f(x_1, \cdots, x_n)$ be a function defined on a neighborhood of a closed set S in euclidean n-space. Suppose that the points of S at which $f_{x_i} = 0$ are interior to S and that the determinant $|f_{x_ix_k}|$ is different from zero at these points. A point at which $f_{x_i} = 0$ is called a critical point and its index is defined to be the negative index of the quadratic form $f_{x_ix_k}\xi_i\xi_k$ at the point. Let m_k be the number of critical points on S of index k and let r_k be the k-th connectivity of S. Then the numbers $h_k = m_k - r_k$ satisfy the relations

$$h_k \geqq h_{k-1} - h_{k-2} + \cdots + (-1)^{k-1}h_0 \geqq 0 \tag{1.1}$$

if the outer normal derivative of $f(x)$ on the boundary of S is positive. This result, which is due to Morse (III),[1] has been generalized by Morse (XVII) to functions on a metric space and, in particular, to functions of curves appearing in the simple problems of the calculus of variations. The special cases where $k \leqq 1$ were treated earlier by G. D. Birkhoff (II) and Poincaré (I).

In the earlier papers the relations (1.1) were obtained by observing the variation of the connectivities of the sets $f \leqq c$ as c increased through critical values. More recently, a second method of approach has been used. This method is based essentially on a maximum-minimum principle and seems to have its beginning with G. D. Birkhoff in 1927 in his book on Dynamical Systems (V, p. 135). Here Birkhoff establishes the existence of closed extremals of index $k > 0$ with the help of a concept which, in the terminology of Morse, is the inferior limit of a k-cycle. Birkhoff treated only a special case. Later Morse (X, p 272) showed that the relations $m_k \geqq r_k$ could be obtained by the use of inferior limits of absolute k-cycles. Birkhoff and Hestenes (XI) then pointed out that the complete relations (1.1) could be obtained if one used the inferior limits of modular k-cycles as well as those of absolute k-cycles. They did not use the term "inferior limit" but used an equivalent concept.

* Received June 8, 1940; Revised January 25, 1942. Presented to the American Mathematical Society April 11, 1936, under the title "Minimax principle in the Calculus of variations"; April 9, 1938, under the title "A theory of critical points of functions in the calculus of variations."

[1] Roman numerals in parentheses refer to the list of papers at the end of this paper. Further references on the theory of critical points are found in these papers.

Moreover deformations were used where homologies should have been used. Later Morse (XII, XVII) introduced the notion of superior limit of an absolute k-cycle. One readily sees that the superior limit of an absolute k-cycle, if finite, is in general the inferior limit of a modular $(k+1)$-cycle. By the use of superior and inferior limits of cycles Morse has developed an elaborate theory of critical points of functions on a metric space. An interesting book on the theories of Morse has been written recently by Seifert and Threlfall (XVIII). References to the various writers on this subject are found in the papers of Morse.

In the present paper a theory of critical points is developed on a metric space which is quite different from that given by Morse but is equivalent to it. The method used is basically the one outlined by Birkhoff and Hestenes (XI, XIV), modified in many ways. It consists essentially of a study of the properties of k-cycles u on F mod F', where F and F' are restricted to belong to a given class $\mathfrak{F}$ with prescribed properties. The properties assigned to $\mathfrak{F}$ are ones enjoyed by the sets determined by inequalities of the form $f(x) \leqq c$, where $f(x)$ is a real valued function and c is a constant. It is shown that the topological theory here developed is based only on these properties. In fact even critical points can be defined in a topological way without specifying the functions $f(x)$ used to determine the class $\mathfrak{F}$, at hand. This development is made in Sections **2-11** below. It is pointed out in Section **12** that our assumptions are not the most general ones that could be used and an indication is given how this extension can be carried out. The extension is not developed in detail in the present paper because in the first place the theory here developed is sufficiently general for our purposes and in the second place it is simpler not to do so.

The paper is divided into two parts; the first is concerned with the topological aspects of the theory and the second with an application to the calculus of variations. Two topological definitions of critical points are given, the first by means of homologies (**5**) and the second by means of homotopic deformations (**7**) in a manner similar to that given by Morse. Analogues of the critical point relations are given in **10** and **11** and applications to functions of n real variables are given in **13** and **14**.

I. TOPOLOGICAL CONSIDERATIONS.

2. Vietoris chains on a metric space S. The topological part of the present paper is based on the concept of Vietoris chains. It is the purpose of the present section to define Vietoris chains and to recall the properties of

these chains that will be useful in the present paper. In general, proofs of these properties will be omitted. The definition of Vietoris cycles here given is somewhat different from the one usually given but is equivalent to it. The author has found this definition to be a very convenient one. It emphasizes a property of k-cycles that is perhaps the most essential property, namely, that in an homology relation any sequence of vertex k-chains defining a Vietoris k-cycle can be replaced by a subsequence.

Vietoris chains are defined in terms of algebraic or vertex chains and these in turn are defined in terms of simplices. By an *oriented k-simplex z on a metric space S* will be meant a class composed of an ordered set P of $k+1$ points on S and all ordered sets obtained from P by an even permutation of its points. If two or more points of P are coincident, the simplex will be said to be *degenerate*. To every set of $k+1$ points $(k>0)$ on S there correspond two simplices, one of which will be taken to be positively oriented and the other negatively oriented. They will be denoted respectively by the symbols z and $-z$. If the simplex is degenerate, z and $-z$ are identical. A single point on S will be said to define a positively oriented 0-simplex.

Let Λ be an additive abelian group of elements λ. *Algebraic* or *vertex k-chains* over Λ are defined in the usual manner as the formal sum

$$u = \Sigma\lambda_i z_i = \lambda_1 z_1 + \cdots + \lambda_m z_m, \tag{2.1}$$

where $\lambda_1, \cdots, \lambda_m$ are elements in Λ and the z's are positively oriented k-simplices, not necessarily distinct. We identify with u the k-chain obtained from u by replacing the sum $\lambda_i z_i + \lambda_j z_j$ by $(\lambda_i + \lambda_j) z_i$ when $z_i = z_j$ and the k-chain obtained from u by deleting a term $\lambda_i z_i$ in which $\lambda_i = 0$ or z_i is degenerate. A vertex k-chain therefore can be represented uniquely in the form (2.1) in which the simplices are non-degenerate, distinct and have non-zero coefficients. Such a form will be called the *reduced form* of u. Two k-chains are identical if they have the same reduced form. The vertices of the k-simplexes in the reduced form of u will be called *vertices* of u and the maximum of the diameters of the simplices in the reduced form will be called the *norm* of u. The boundary u' of u is defined in the usual manner. We have $(u')' = 0$. By the sum $u + v$ of the k-chain (2.1) and $v = \Sigma\lambda^*_j z^*_j$ will be meant the k-chain $\Sigma\lambda_i z_i + \Sigma\lambda^*_j z^*_j$. It is clear that $(u+v)' = u' + v'$.

An infinite sequence (u_n) of vertex k-chains, whose norms converge to zero as n becomes infinite, will be called a *Vietoris k-chain u on S* if the closure of the point set composed of the vertices of the vertex k-chains $u_1, u_2, \cdots$ forms a compact set U. The points in U will be called *vertices* of u. If U is null, we say that u is null and write $u = 0$. Let $u = (u_n)$, $v = (v_n)$ be two

Vietoris k-chains. The sequences $(u_n + v_n)$ and $(u_n - v_n)$ define Vietoris k-chains which we shall call the sum $u + v$ and the difference $u - v$ of u and v. We say that $u = v$ in case $u - v = 0$. By the *boundary* u' of a k-chain $u = (u_n)$ will be meant the $(k-1)$-chain defined by the sequence (u'_n), where u'_n is the boundary of u_n. We have the relations $(u')' = 0$, $(u + v)' = u' + v'$.

A k-chain u will be said to be *on a subset* F of S if its vertices are on F. The null k-chain is on every subset of S and, in particular, is on the null set. Let u be a k-chain on F and let F' be a subset of F such that u' is on F'. We shall say that the k-chain u is *homologous to zero* on F mod F', written $u \sim 0$ on F mod F', if there is a $(k+1)$-chain w on F such that the k-chain $w' - u$ is on F'. Two k-chains u and v on F will be said to be *homologous on* F *mod* F', written $u \sim v$ on F mod F', if $u - v \sim 0$ on F mod F'. When $F = S$ the phrase "on F" will be omitted. Likewise when $F' = 0$ the phrase "mod F'" will be omitted.

By a *subchain* v of a k-chain $u = (u_n)$ will be meant a k-chain defined by an infinite subsequence (v_n) of (u_n). Consider now two sets F, F' on S such that $F \supset F'$. By a *k-cycle* u *on* F *mod* F' will be meant a k-chain u on F whose boundary u' is on F' and which is such that the relation, $u \sim v$ on F mod F', holds for every subchain v of u. It is easily seen that if u is a k-cycle on F mod F' and v is any k-chain satisfying the relation, $u \sim v$ on F mod F', then v is also a k-cycle on F mod F'. In particular, every subchain of a k-cycle on F mod F' is a k-cycle on F mod F'. The k-cycles here defined are Vietoris k-cycles, as can be seen by interpreting the result described in the following lemma in terms of e-homologies.

LEMMA 2.1. *Let $F \supset F'$ be subsets of S and let $u = (u_n)$ be a k-chain on F whose boundary u' is on F'. Then u is a k-cycle on F mod F' if and only if it is homologous on F mod F' to the subchain v of u defined by the sequence (u_{n+1}).*

It is clear from the definition of k-cycles that if u is a k-cycle on F mod F', then $u \sim v$ on F mod F', where $v = (u_{n+1})$. Conversely, suppose that this homology relation holds. Then there is a $(k+1)$-chain $w = (w_n)$ on F such that the k-chain $w' - u + v$ is on F', and hence such that the vertex k-chains $w'_n - u_n + v_n$ $(n = 1, 2, \cdots)$ are on F'. Consider now a subchain $\bar{v} = (u_{n_i})$ of $u = (u_n)$. Set $\bar{w}_i = 0$ if $n_i = i$ and set $\bar{w}_i = w_i + w_{i+1} + \cdots + w_{n_i - 1}$ otherwise. Then $\bar{w}_i - u_i + u_{n_i} = 0$ when $n_i = i$ and when $n_i > i$ the vertex k-chain

$$\bar{w}_i - u_i + u_{n_i} = \sum_{j=i}^{j=n_i-1} (w'_j - u_j + u_{j+1}),$$

is on F'. Setting $\bar{w} = (\bar{w}_n)$ it follows that the Vietoris k-chain $\bar{w}' - u + \bar{v}$ is on F' and hence that $u \sim \bar{v}$ on $F \bmod F'$. This completes the proof of Lemma 2.1.

In order to insure that our k-cycles have the properties described in the next two lemmas, we make the following

ASSUMPTION. *Let F and $F' \subset F$ be compact sets and let k be a positive integer. Given a constant $e > 0$ there is a positive constant $e' < e$ such that if u is a vertex k-chain on F of norm at most e' with u' on F' and e'' is a positive constant, there exist on F a vertex k-chain v of norm at most e'' and a $(k+1)$-chain w of norm at most e, such that $w' - u + v$ and v' are on F'.*

This assumption holds for the cases in which we are interested. In particular it holds when Λ is a field.

The following lemma will be used frequently.

LEMMA 2.2. *Let u be a k-cycle on $F \bmod F'$ and let G be a subset of u such that $u \sim 0$ on $F \bmod F' + G$. There is a k-cycle v on $F' + G \bmod F'$ such that $u \sim v$ on $F \bmod F'$. In fact, if F' and G are closed, there is a k-cycle v on $G \bmod F'G$ such that $u \sim v$ on $F \bmod F'$.*

In the proof one can assume, without loss of generality, that the sets $F, F', F' + G$ are compact. The lemma can be established by an argument like that used by Morse (XVII, Theorem 3.1, p. 396) for the case in which Λ is a field. In a similar manner (cf. Morse XVII, pp. 396, 392, Theorems 3.2 and 3.3 and Lemma 2.2), one obtains the following:

LEMMA 2.3. *Let $F \supset F'$ and G be closed subsets of S and let u be a k-cycle on $F \bmod F'$.*

1. *If $G \supset F$ and $u \sim 0$ on $G \bmod F'$, there is a $(k+1)$-cycle w on $G \bmod F$ such that $u \sim w'$ on $F \bmod F'$.*

2. *If $G \subset F'$ and $u' \sim 0$ on $F' \bmod G$, where u' is the boundary of u, there is a k-cycle v on $F \bmod G$ such that $u \sim v$ on $F \bmod F'$.*

3. *If F and F' are compact and $u \sim 0$ on $N \bmod N'$ for every pair of neighborhoods N of F and N' of F', then $u \sim 0$ on $F \bmod F'$.*

3. Admissible classes $\mathfrak{F}$ of sets F on S. Consider for the moment a real single valued lower semicontinuous function $f(x)$ on S and let $\mathfrak{F}$ be the class of all sets F on S determined by an inequality of the form $f(x) \leqq c$, where c is a constant or $\pm \infty$. The topological part of the theory of critical points of $f(x)$ on S here developed is essentially a theory of k-cycles on $F \bmod F'$,

where F and F' belong to $\mathfrak{F}$. It is based on the following properties of $\mathfrak{F}$: (1) *Given two sets in* $\mathfrak{F}$, *one is a subset of the other*; (2) *the intersection of the sets belonging to a subclass of* $\mathfrak{F}$ *is in* $\mathfrak{F}$; (3) *the sets in* $\mathfrak{F}$ *are closed*; (4) *the null set* $F = 0$ *and the set* $F = S$ *are in* $\mathfrak{F}$. In order to emphasize the dependence of our theory on these properties, we shall begin with an arbitrary class $\mathfrak{F}$ having these properties. Such a class will be called *admissible.* The class $\mathfrak{F}$ related to a function $f(x)$ as described above will be called the *admissible class determined by* $f(x)$. Since the results described below depend only on the four properties listed above, it follows that two functions $f(x)$ and $g(x)$, which determine the same admissible class $\mathfrak{F}$, give rise to the same topological theory. They have accordingly the same critical k-sets (**5**), the same homotopic critical points (**7**), etc. Moreover, they are connected by a relation of the form $f(x) = h[g(x)]$, where $h(t)$ is a non-decreasing function of a real variable t.

Consider now a particular admissible class $\mathfrak{F}$. In the sequel the symbol (u, F, F') will be used to denote a k-cycle u on F mod F' for which F, F' are in $\mathfrak{F}$ and $F \supset F'$. Two k-cycles (u, F, F') and (v, G, G') will be said to be *homologous,* written $(u, F, F') \sim (v, G, G')$, if $u \sim v$ on $F + G$ mod $F' + G'$. It is clear that either $F + G = F$ or $F + G = G$ in view of property (1) for the admissible class $\mathfrak{F}$. A k-cycle (u, F, F') will be said to be *homologous to zero,* written $(u, F, F') \sim 0$, if $u \sim 0$ on F mod F'.

We shall center our attention on special k-cycles called minimizing k-cycles. By a *minimizing k-cycle* (u, F, F') will be meant one such that the relation $F \subset G$ holds for every k-cycle (v, G, G') that is homologous to it. In view of this definition, the null k-cycle $(0, 0, 0)$ is minimizing. This definition is a simple modification of one used by Birkhoff and Hestenes (XI, XIV) in a similar situation. In the terminology of Morse (XVII, p. 408), the k-chain u belonging to a non-null minimizing k-cycle (u, F, F') would be called a k-cap.

LEMMA 3.1. *A k-cycle* (u, F, F') *is minimizing if and only if there is no proper subset* F'' *of* F *in* $\mathfrak{F}$ *such that* $u \sim 0$ *on* F *mod* $F' + F''$.

This criterion is obviously satisfied when $F = 0$. Suppose therefore that $F \neq 0$. If $(u, F, F') \sim 0$, the relation $u \sim 0$ on F mod $F' + F''$ holds with $F'' = 0$. In this case the k-cycle (u, F, F') is not minimizing. We can assume, therefore, that $(u, F, F') \not\sim 0$. Then $F' \neq F$. If there is a proper subset F'' of F in $\mathfrak{F}$ such that $u \sim 0$ on F mod $F' + F''$, then (u, F, F') is homologous to a k-cycle of the form $(v, F'', F'F'')$, by Lemma 2.2 with $G = F''$. Since F'' is a proper subset of F, the k-cycle (u, F, F') cannot be minimizing

in this event. Finally, suppose that the k-cycle (u, F, F') is not minimizing and let (v, G, G') be a k-cycle homologous to it such that $FG \neq F$. Then G is a proper subset of F and the relation $u \sim 0$ on $F \bmod F' + G$ holds, since $u \sim v$ on $F \bmod F' + G'$. The homology relation described in the lemma accordingly holds with $F'' = G$. This completes the proof of the lemma.

COROLLARY. *If (u, F, F') is a minimizing k-cycle and G is a set in $\mathfrak{F}$ such that u is on G, then $F \subset G$.*

For since u is on F and on G it is also on FG. Hence $u \sim 0$ on $F \bmod F' + FG$. By the last lemma with $F'' = FG$, it follows that $FG = F$, that is, $F \subset G$, as was to be proved.

If $\mathfrak{F}$ is the class of sets related to a function $f(x)$, as described above, the corollary states that for a minimizing k-cycle (u, F, F') the least upper bounds of $f(x)$ on F and on the set of vertices of u are the same.

In order to facilitate generalizations in **12** below and to avoid excessive repetitions of the phrase "a proper subset of F in $\mathfrak{F}$" we shall define a proper subset of F in $\mathfrak{F}$ to be a *primary subset* of F. With the definition in mind we can prove the following:

LEMMA 3.2. *Let (u, F, F') be a non-null minimizing k-cycle. The set F' is a primary subset of F. If (v, G, G') is a k-cycle homologous to (u, F, F'), it is minimizing if and only if $F = G$ and G' is a primary subset of F. The k-cycle $(u, F, F' + F'')$ is minimizing for every primary subset F'' of F.*

The set F' is a primary subset of F, by virtue of the criterion described in the last lemma. Consider now a k-cycle (v, G, G') that is homologous to (u, F, F'). If it is minimizing, then obviously $F = G$ and G' is a primary subset of F. Conversely, if these relations hold, the k-cycle (v, G, G') must be minimizing. Otherwise there would exist, by Lemma 3.1, a proper subset G'' of G in $\mathfrak{F}$ such that $v \sim 0$ on $F \bmod G' + G''$. Moreover the set $F'' = G' + G''$ would be a proper subset of $F = G$, since G' is a primary subset of F. From the relation $u \sim v$ on $F \bmod F' + G'$ it would follow that the relation $u \sim 0$ on $F \bmod F' + F''$ would hold with $F'' = G' + G''$. This is impossible, by Lemma 3.1. Hence (v, G, G') is minimizing in this case. The last statement in the lemma follows from the preceding one and the fact that the sum of two primary subsets of F is a primary subset of F.

LEMMA 3.3. *Two minimizing k-cycles (u, F, F') and (v, G, G'), not both null, are homologous to a third minimizing k-cycle if and only if $F = G$ and there is a primary subset F'' of F such that $u \sim v$ on $F \bmod F''$.*

For let (w, H, H') be a minimizing k-cycle that is homologous to two minimizing k-cycles (u, F, F') and (v, G, G'). Then $F = H = G$ and $u \sim v$ on $F \bmod F''$ with $F'' = F' + G' + H'$. If $F \neq 0$, the set F'' is a primary subset of F, by the last lemma. Conversely if $u \sim v$ on $F \bmod F''$, where F'' is a primary subset of F, then (u, F, F') and (v, G, G') are homologous to the minimizing k-cycle $(u, F, F' + G' + F'')$. This proves the lemma.

4. Existence of minimizing k-cycles. As a first existence theorem we have

THEOREM 4.1. *If the sets $F \neq S$ in $\mathfrak{F}$ are compact, every k-cycle (u, F, F') is homologous to a minimizing k-cycle (v, G, G') having $G' \subset F'$. In fact the k-cycle (v, G, G') can be chosen so that the $(k-1)$-cycle $(v', G', 0)$ is minimizing, where v' is the boundary of v.*

In order to prove the first statement in the theorem, consider a k-cycle (u, F, F') having $F \neq S$. If $(u, F, F') \sim 0$, choose $(v, G, G') = (0, 0, 0)$. Suppose, therefore, that $(u, F, F') \not\sim 0$. Let $\mathfrak{F}''$ be the class of all subsets F'' of F in $\mathfrak{F}$ such that the relation $u \sim 0$ on $F \bmod F' + F''$ holds. Let G be the intersection of all sets in $\mathfrak{F}''$. Then G is in $\mathfrak{F}$, by property (2) of $\mathfrak{F}$. In fact G is in $\mathfrak{F}''$. To prove this result, observe that on every neighborhood N' of G there is a set F'' belonging to $\mathfrak{F}''$. Consequently $u \sim 0$ on $F \bmod F' + N'$ for every neighborhood N' of G. By the use of part 3 of Lemma 2.3 it is seen that $u \sim 0$ on $F \bmod F' + G$ and hence that G is in $\mathfrak{F}''$. From Lemma 2.2 it follows that there is a k-cycle of the form $(v, G, F'G)$ that is homologous to (u, F, F'). We shall show that $(v, G, F'G)$ is minimizing. If it were not so there would exist, by Lemma 3.1, a proper subset G'' of G in $\mathfrak{F}$ such that $v \sim 0$ on $G \bmod F'G + G''$. Hence $u \sim 0$ on $F \bmod F' + G''$ since (u, F, F') and $(v, G, F'G)$ are homologous. But this would imply that G'' is in $\mathfrak{F}''$ contrary to our choice of G as the intersection of the sets in $\mathfrak{F}''$. The k-cycle $(v, G, F'G)$ is accordingly minimizing and the first statement in the theorem is proved for the case in which $F \neq S$. Suppose now that $F = S$ and that (u, F, F') is not minimizing. Then there is a k-cycle (w, H, H') with $H \neq F$ and $H' \subset F'$ that is homologous to (u, F, F'). We have just seen that there is a minimizing k-cycle (v, G, G') homologous to (w, H, H') such that $G' \subset H'$. It is clear that $(u, F, F') \sim (v, G, G')$. This completes the proof of the first part of the theorem.

In order to prove the last statement in the theorem select a minimizing k-cycle (v, G, G') homologous to (u, F, F') with $G' \subset F'$. If the $(k-1)$-cycle $(v', G', 0)$ is not minimizing, there is a proper subset F'' of G' in $\mathfrak{F}$ such

that $v' \sim 0$ on G' mod F''. Let $\mathfrak{F}''$ be the class of all sets F'' of this type and let G'' be the intersection of the sets in $\mathfrak{F}''$. Then, as in the last paragraph, $v' \sim 0$ on G' mod G'', since G' and G'' are compact. By virtue of part 2 of Lemma 2. 3 there is a k-cycle (w, G, G'') that is homologous to (v, G, G') and hence to (u, F, F'). From our choice of the set G'' it follows that the $(k-1)$-cycle $(w', G'', 0)$ is minimizing. The k-cycle (w, G, G''), being homologous to (v, G, G'), is minimizing. This completes the proof of Theorem 4. 1.

THEOREM 4. 2. *Suppose that the sets $F \neq S$ in $\mathfrak{F}$ are compact, and let (u, F, F') be a non-null minimizing k-cycle such that $u \sim 0$ on S mod F'. There exists a minimizing $(k+1)$-cycle (v, G, F) such that $(v', F, 0) \sim (u, F, F')$ where v' is the boundary of v. In fact (v, G, F) can be chosen so that, given a proper subset G'' of G in $\mathfrak{F}$, one has $u \not\sim 0$ on G'' mod F'.*

For let (u, F, F') be a minimizing k-cycle such that $u \sim 0$ on S mod F' and let $\mathfrak{F}''$ be the maximal subclass of $\mathfrak{F}$ such that $u \sim 0$ on F'' mod F' for every set F'' in $\mathfrak{F}''$. Then the intersection G of the sets in $\mathfrak{F}''$ has the property that $u \sim 0$ on G mod F', by virtue of part 3 of Lemma 2. 3 and the compactness of the sets $F \neq S$ in $\mathfrak{F}$. Using part 2 of the same lemma one finds that there is a $(k+1)$-cycle (v, G, F) such that $u \sim v'$ on F mod F', that is, such that $(u, F, F') \sim (v', F, 0)$. This proves the theorem.

THEOREM 4. 3. *The conclusions described in Theorems* 4. 1 *and* 4. 2 *hold true if $\mathfrak{F}$ has the property that given a set $F \neq S$ in $\mathfrak{F}$ and a compact subset C of F there is a compact subset G of F in $\mathfrak{F}$ such that $C \subset G$.*

The proof is like that given above. One uses the additional fact that if u is a k-cycle on F mod F' there exist compact subsets G, G' of F, F' respectively such that u is a k-cycle on G mod G'.

The hypotheses in Theorem 4. 3 are satisfied, for example, when the class $\mathfrak{F}$ is the class of all sets determined by inequalities of the form $f(x) \leqq c$ or $f(x) < c$, where $f(x)$ is continuous on S, provided the sets of the form $f(x) \leqq c$ are compact. In this case the assumption of closure of the sets in F is not satisfied. This assumption is not needed in the proof of Theorem 4. 3.

5. Critical k-sets. In the introduction critical points of index k were defined for a function of n real variables. It is the purpose of the present section to define analogues of these points by means of homologies. The analogue of a critical point of index k here given is, in general, not a single point but a set of points, which we shall call a critical k-set. It will be seen in **14** below that a critical point of index k forms a critical k-set.

Consider now a minimizing k-cycle (u, F, F'). In the study of the

critical k-sets defined below, two special types of subsets of F play a basic role. These types will be designated by A and B, respectively. Briefly a set X is of type A if the minimizing properties of (u, F, F') are unaltered when the points exterior to a neighborhood of X are disregarded. More precisely, by a *subset of F of type A relative to* (u, F, F') will be meant a closed subset X of F with the property that, given a neighborhood N of X, there is no primary subset F'' of F such that $u \sim 0$ on $F \bmod F'' + V$, where $V = F - FN$. A subset X of F of type A relative to (u, F, F') will be said to be of *type B relative to* (u, F, F') if the set XX' is of type A for every set X' of type A. The phrase "relative to (u, F, F')" will be omitted when no confusion arises by so doing. The set $X = F$ is of type B. Moreover if X' and X'' are subsets of F of type B so also is their intersection $X'X''$. For since X' and X'' are of type B, it is seen successively that the sets XX' and $XX'X''$ are of type A whenever X is of type A. Consequently $X'X''$ is of type B, as was to be proved.

By the *critical k-set C determined by a minimizing k-cycle* (u, F, F') will be meant the intersection of all subsets X of F of type B relative to (u, F, F'). It will be seen in Theorem 5.1 below that the set $X = C$ is a subset of F of type B and hence that the set $X = C$ is the smallest subset of F of type B. In general, the critical k-set C is not a member of the admissible class $\mathfrak{F}$ by which it is determined.

In order to understand more clearly the significance of these special subsets and of the definition of critical k-sets, consider the following simple example. Let S be the interval $0 \leqq x \leqq 3\pi$ and let $\mathfrak{F}$ be the admissible class of sets determined by all inequalities of the form $\sin x \leqq c$, where c is a constant. Denote by F and F' the sets in $\mathfrak{F}$ determined by the inequalities $\sin x \leqq 1$ and $\sin x \leqq 0$, respectively. Using the integers modulo 2 as coefficients, let u be the 1-chain defined by the interval $0 \leqq x \leqq 3\pi$ with $x = 0$ and $x = 3\pi$ as its boundary. The 1-cycle (u, F, F') is then minimizing. A closed subset X of F is of type A if, and only if, it contains one of the points $x' = \pi/2$, $x'' = 5\pi/2$ at which $\sin x = 1$. It is of type B if and only if it contains both x' and x''. The set C composed of x' and x'' is the smallest subset of F of type B and is accordingly the critical 1-set determined by (u, F, F').

The following lemma will be useful.

LEMMA 5.1. *Let (u, F, F') be a non-null minimizing k-cycle and let X be a subset of F of type A relative to (u, F, F'). If V denotes the set of vertices of the k-chain v belonging to a k-cycle (v, F, F'') homologous to (u, F, F'), there is no neighborhood N of X such that the set NV is on a primary subset of F. Moreover the set XV is non-null.*

For suppose that there exists a neighborhood N of X and a primary subset G of F such that $NV \subset G$. Then $V \subset G + Y$, where $Y = F - FN$. Since the sum $F' + F'' + G$ of primary subsets of F is also a primary subset of F, we can suppose that this sum is identical with G. Then $G \supset F' + F''$. From the relations, $u \sim v$ on F mod $F' + F''$ and $V \subset G + Y$, it follows that $u \sim 0$ on F mod $G + Y$, contrary to our assumption that X is of type A. This proves the first part of the lemma. The second follows from the first, since if XV were null, the set NV would be null for a sufficiently small neighborhood N of X.

THEOREM 5. 1. *The critical k-set C determined by a non-null minimizing k-cycle (u, F, F') is a non-null subset of F of type B relative to (u, F, F'). If V denotes the set of vertices of the k-chain v belonging to a minimizing k-cycle (v, F, F'') homologous to (u, F, F'), the set CV is non-null. In fact, the closure C' of the sum of all sets CV of this type is identical with C.*

In order to show that the set CV described in the theorem is non-null, observe that, by virtue of Lemma 5. 1, the set XV is non-null for every subset X of F of type A. The set XV is compact since V is compact. Since the class of subsets of F of type B is closed relative to intersections, the intersection CV of all sets of the form XV, where X is of type B, is non-null, as was to be proved. The set C is closed since it is the intersection of closed sets.

The remaining statements of the theorem will be established by showing that the set $X = C'$ is of type B. For then C and C', being subsets of each other, must be identical. Suppose, therefore, that the set C' is not of type B. Then there exists a subset X of F of type A such that the set XC' is not of type A. We may accordingly choose a neighborhood N of XC' such that $u \sim 0$ on F mod $F'' + Y$, where $Y = F - FN$. By virtue of Lemma 2. 2 we can suppose that the set U of vertices of u is on $F'' + Y$. The set NU is then on a primary subset of F. Moreover, $C'U = CU$ in view of our definition of C'. Since U is compact and the set $XC'U = XCU$ is the product of all sets of the form $XX'U$, where X' is a subset of F of type B, there is a subset X' of F of type B such that $XX'U \subset NU \subset F''$. The set $XX'U$ is, therefore, on the primary subset F'' of F, which is impossible, by Lemma 5. 1, since the set XX' is of type A, in view of the fact that X is of type A and X' of type B. Our assumption that the set C' is not of type B is false and the theorem is established.

COROLLARY 1. *If C is the critical k-set determined by a minimizing k-cycle (u, F, F') and $F' \subset E = F - C$, then $u \not\sim 0$ on F mod E. The relation $F' \subset E$ holds if no primary subset of F has a point in common with C.*

For if $u \sim 0$ on $F \bmod E$ one could suppose, by Lemma 2.2, that u is on $F' + E = E$ and hence would have no vertices in C, contrary to the second statement in Theorem 5.1.

COROLLARY 2. *Homologous minimizing k-cycles determine the same critical k-sets. In fact, if two minimizing k-cycles are homologous to the same minimizing k-cycle, they determine the same critical k-set.*

The critical k-set C belonging to a minimizing k-cycle (u, F, F') is determined completely by the k-cycles on a neighborhood of C, as is shown in the following:

THEOREM 5.2. *Let N be a neighborhood of the critical k-set determined by a non-null minimizing k-cycle (u, F, F'). There exists a primary subset F'' of F such that $u \sim 0$ on $F \bmod F'' + FN$. In fact, there exists a k-cycle v on $FN \bmod F''N$ such that the system (v, F, F'') defines a minimizing k-cycle homologous to (u, F, F'). If V denotes the set of vertices of a k-chain v of this type, the closure C' of the sum of all sets of the form CV is identical with C.*

In order to prove the first two conclusions in the theorem observe that the set $X = F - FN$, having no points in common with C, cannot be of type A. There exists accordingly a neighborhood N' of X and a primary subset F'' of F such that $u \sim 0$ on $F \bmod F'' + Y$, where $Y = F - FN'$. By Lemma 2.2 there exists a k-cycle v on $Y \bmod F''Y$ such that $u \sim v$ on $F \bmod F''$. Since $Y \subset FN$ the k-chain v is a k-cycle on $FN \bmod F''N$. Moreover, the system (v, F, F'') is a k-cycle homologous to (u, F, F') and, by Lemma 3.2, it is minimizing. In view of the fact that v is on FN and $u \sim v$ on $F \bmod F''$ it follows that $u \sim 0$ on $F \bmod F'' + FN$.

It remains to prove the last statement in the theorem. As in the proof of Theorem 5.1 it is sufficient to show that the set C' described in Theorem 5.2 is a subset of F of type B. Suppose that the contrary is true. Then there is a subset X of F of type A such that the set CX' is not of type A. Let N' be a neighborhood of XC' and let G be a primary subset of F such that $u \sim 0$ on $F \bmod G + Y'$, where $Y' = F - FN'$. By Lemma 2.2 we can suppose that the set U of vertices of u is on $G + Y'$. The remainder of the proof is like that of Theorem 5.1 provided that we can show that the k-chain $u = (u_n)$ can be chosen so that u is a k-cycle on $FN \bmod F''N$, where F'' is a primary subset of F. To this end let $v = (v_n)$ be a k-chain related to u as described in the last paragraph. Then there is a primary subset F'' of F such that v is a k-cycle on $FN \bmod F''N$ and $u \sim v$ on $F \bmod F''$. We can suppose that

$F'' \supset F' + G$. Let $w = (w_n)$ be a $(k+1)$-chain on F such that the k-chain $w' - u + v$ is on F'' and select a neighborhood N'' of C' such that the closures of N' and N'' lie respectively in N'' and N. From the definition of k-cycles the sequence $u = (u_n)$ can be replaced by any infinite subsequence. Consequently we can suppose that every simplex in the vertex chains u_n, v_n, w_n $(n = 1, 2, \cdots)$ having a vertex in N'' is in N and that every simplex having a vertex in N' is in N''. Let $\bar{u} = (\bar{u}_n)$ be the k-chain obtained from $u = (u_n)$ by deleting the k-simplices in u_n that are exterior to N''. Similarly let $\bar{w} = (\bar{w}_n)$ be the $(k+1)$-chain obtained from $w = (w_n)$ by deleting the $(k+1)$-simplices exterior to N''. The chains $\bar{u}$, $\bar{w}$ are on FN, are identical with u, w on FN' and are such that the k-chain $\bar{w}' - \bar{u} + v$ is on $F'' + E$, where $E = F(N - N'')$. Let $\overline{v} = (\overline{v}_n)$ be the k-chain obtained from $\bar{w}' - \bar{u} + v$ be deleting all k-simplices of $\bar{w}'_n - \bar{u}_n + v_n$ having no vertex in E. The k-chain $\overline{v}$ has no vertex in N' since, by construction, no simplex of the vertex k-chains u_n, v_n, w_n has vertices in N' and in E. It follows that the vertices of $\bar{u} + \overline{v}$ in N' are vertices of $\bar{u}$ and hence of u. Since u is on $G + Y'$ and $F'' \supset G$ it follows that $\bar{u} + \overline{v}$ is on $F''N + Y'N$. Moreover, the k-chain $\bar{w}' - (\bar{u} + \overline{v}) + v$ is on $F''N$. Hence $\bar{u} + \overline{v} \sim v$ on $FN \bmod F''N$. The k-chain $\bar{u} + \overline{v}$ is accordingly a k-cycle on $FN \bmod F''N$ since v has this property. It is clear that $u \sim \bar{u} + \overline{v}$ on $F \bmod F'$. Replacing u by $\bar{u} + \overline{v}$ we obtain the desired result, namely, that u can be chosen to be on a set $F'' + Y$ in such a way that u is a k-cycle on $FN \bmod F''N$ and F'' is a primary subset of F. This completes the proof of Theorem 5. 2.

THEOREM 5. 3. *No neighborhood N of a point x in the critical k-set C determined by a minimizing k-cycle (u, F, F') is such that the set FN is on a primary subset F'' of F.*

This result follows from Theorems 7. 1 and 7. 2 below. A direct proof can be made as follows. Let x be a point of F for which there exists a neighborhood N of x such that FN is on a primary subset F'' of F. If X is a subset of F of type B, it is easily seen that the closure X' of the set $X - F''X$ is also of type B so that $C \subset X'$. Taking $X = C$ one has $X' = C$. The point x therefore cannot be in C and the theorem is proved.

As an application of the last theorem we have the following:

COROLLARY. *Let $\mathfrak{F}$ be the admissible class of sets $f(x) \leqq c$ determined by a real single valued lower semi-continuous function $f(x)$ on S, as described in* **3**. *Let C be the critical k-set determined by a non-null minimizing k-cycle (u, F, F') and denote by b the least upper bound of $f(x)$ on F. The upper*

limit of $f(x)$ on F at a point x' in C is equal to b. If $f(x)$ is continuous, then $f(x) = b$ on C. If $f(x) = b$ on C, then $F' \subset E = F - C$ and $u \not\sim 0$ on F mod E.

The first conclusion in the corollary follows at once from Theorem 5.3 since if the upper limit of $f(x)$ on F at a point x' in C were not equal to b there would exist a neighborhood N of x' and a constant $c < b$ such that FN would be on the primary subset F'' of F determined by the inequality $f(x) \leqq c$. The last statement in the lemma follows from Corollary 1 to Theorem 5.1.

It is desirable to obtain simple criteria for a point to form a critical k-set. To this end we have the following:

THEOREM 5.4. *A point x' forms a critical k-set if and only if there exists a k-cycle (u, F, F') such that given a neighborhood N of x' then* (1) *there is a primary subset F'' of F such that $u \sim 0$ on F mod $F'' + FN$;* (2) *there is no primary subset F'' of F such that $u \sim 0$ on F mod $F' + F'' + Y$, where $Y = F - FN$.*

The necessity of condition (1) follows from Theorem 5.2 with $C = x'$. The condition (2) is the condition that the set $C = x'$ be of type A and hence also is a necessary condition. To prove the converse observe that when condition (2) holds the k-cycle (u, F, F') must be minimizing, by Lemma 3.1. This condition also implies that any closed subset X of F containing x' is of type A. On the other hand every subset X of F of type A contains x'. If this were not so, one could select a neighborhood N of x' and a neighborhood N' of X having no points in common. By condition (1) it would follow that there exists a primary subset F'' of F such that $u \sim 0$ on F mod $F'' + FN$ and hence such that $u \sim 0$ on F mod $F'' + Y'$, where $Y' = F - FN'$. This would contradict our choice of X as being of type A. Hence $X \supset x'$ and the set $C = x'$ is of type B. It is accordingly the critical k-set determined by (u, F, F') and the theorem is established.

COROLLARY. *In order that a point x' form a critical k-set it is sufficient that there exist a k-cycle (u, F, F') such that* (1) *given a neighborhood N of x' there is a primary subset F'' of F such that $u \sim 0$ on F mod $F'' + FN$;* (2) *there exists no primary subset F'' of F such that x' is on $F' + F''$;* (3) *$u \not\sim 0$ on F mod E, where E is the set obtained from F by deleting the point x'. These conditions are necessary conditions for the point x' to form the critical k-set determined by a minimizing k-cycle (u, F, F') if there is no primary subset of F containing x'.*

The conditions in the corollary obviously imply those in the last theorem.

The last statement in our corollary follows from Corollary 1 to Theorem 1 since our hypotheses imply that $F' \subset E$.

6. Sum of two critical k-sets. Let C_1 and C_2 be the critical k-sets determined by two minimizing k-cycles (u, F, F') and (v, G, G') with $F = G$. If the k-cycle $(u + v, F, F' + G')$ is also minimizing, one might expect that its critical k-set C would be a subset of $C_1 + C_2$. That this is not always the case can be seen by an example. It is true, however, that $C = C_1 + C_2$ when C_1 and C_2 are at a positive distance apart, as is seen in the following:

LEMMA 6.1. *Suppose that the critical k-sets C_1 and C_2, determined by two non-null minimizing k-cycles (u, F, F') and (v, G, G') with $F = G$, are at a positive distance from each other. Then the k-cycle $(u + v, F, F' + G')$ is minimizing and its critical k-set C is identical with $C_1 + C_2$.*

We can suppose, for convenience, that $F' = G'$. Let N_1 and N_2 be neighborhoods of C_1 and C_2, respectively, whose closures have no point in common. By Theorem 5.2 we can suppose that u is on N_1 and v is on N_2. If the k-cycle $(u + v, F, F')$ were not minimizing, there would exist, by Lemma 3.1, a primary subset F'' of F such that $u + v \sim 0$ on $F \bmod F''$. The k-cycles (u, F, F'') and $(-v, F, F'')$ would be homologous minimizing k-cycles determining disjoint critical k-sets C_1 and C_2, contrary to the result described in Corollary 2 of Theorem 5.1. It follows that the k-cycle $(u + v, F, F')$ is minimizing.

As a second step in the proof of Lemma 6.1 it will be shown that a set X is of type A relative to $(u + v, F, F')$ if, and only if, it is of type A relative to (u, F, F') or (v, F, F'). To do so let N be a neighborhood of X and set $Y = F - FN$. If there is a primary subset F'' of F such that $u \sim 0$ on $F \bmod F'' + Y$ and $v \sim 0$ on $F \bmod F'' + Y$, then $u + v \sim 0$ on $F \bmod F'' + Y$. Consequently, if X is not of type A relative to either (u, F, F') or (v, F, F') it cannot be of type A relative to $(u + v, F, F')$. Conversely suppose that X is not of type A relative to $(u + v, F, F')$ but is of type A relative to one of the two other k-cycles, say (u, F, F'). Then since C_1 is of type B relative to (u, F, F'), the set XC_1 is of type A relative to (u, F, F') and we can suppose that X has been replaced by XC_1. Since X is not of type A relative to $(u + v, F, F')$ there is a neighborhood N of X and a primary subset F'' of F such that $u + v \sim 0$ on $F \bmod F'' + Y$, where $Y = F - FN$. Since $X \subset C_1$ we can choose N so that N is interior to the neighborhood N_1 described in the last paragraph. But since v is on Y, it follows that $u \sim 0$ on $F \bmod F'' + Y$, contrary to our assumption that X is of type A relative to

(u, F, F'). This proves the statement made at the beginning of this paragraph.

In order to prove that $C \subset C_1 + C_2$ consider a set X of type A relative to $(u + v, F, F')$. Then X must be of type A relative to either (u, F, F') or (v, F, F'), say the former. Since $C_1 + C_2$ is of type B relative to (u, F, F'), the set $X(C_1 + C_2)$ is of type A relative to (u, F, F') and hence relative to $(u + v, F, F')$. Consequently the set $C_1 + C_2$ is of type B relative to $(u + v, F, F')$ and $C \subset C_1 + C_2$. In order to show that $C \supset C_1$ let X be a set of type A relative to (u, F, F'). Then XC_1 is of type A relative to (u, F, F') and hence also relative to $(u + v, F, F')$. The set XC_1C must accordingly be of type A relative to $(u + v, F, F')$ and hence relative to either (u, F, F') or (v, F, F'). Being in N_1 it must be of type A relative to (u, F, F'), that is, the set C is of type B relative to (u, F, F'). Consequently, $C \supset C_1$. By symmetry $C \supset C_2$ so that $C = C_1 + C_2$. This completes the proof of Lemma 6.1.

LEMMA 6.2. *Let C be the critical k-set determined by a non-null minimizing k-cycle (u, F, F') and suppose that C is the sum of two non-null sets C_1 and C_2 at a positive distance from each other. There exist minimizing k-cycles (v, F, F'') and (w, F, F'') whose critical k-sets are C_1 and C_2, respectively, and are such that $(u, F, F') \sim (v + w, F, F'')$.*

In order to prove this result let N_1 and N_2 be neighborhoods of C_1 and C_2 whose closures have no point in common. Let $N = N_1 + N_2$. By Theorem 5.2 we can suppose that u is a k-cycle on FN mod $F'N$. Then u is the sum of two k-chains $v + w$ such that v is on N_1 and w is on N_2. Since u is a k-cycle on FN mod $F'N$ and N_1, N_2 are disjoint it follows that v is a k-cycle on FN_1 mod $F'N_1$. Similarly w is a k-cycle on FN_2 mod $F'N_2$. Consequently these systems (v, F, F') and (w, F, F') are k-cycles. Suppose that one of them, say the latter, is not minimizing. Then there exists a primary subset F'' of F such that $w \sim 0$ on F mod F''. Let X be a subset of F of type A relative to (u, F, F'). The set C_1, being a proper subset of C, is not of type B relative to (u, F, F'). Consequently we can select X so the set C_1X is not of type A. There is, accordingly, a neighborhood $N' \subset N_1$ of C_1X and a primary subset F''' of F such that $u \sim 0$ on F mod $F''' + Y'$, where $Y' = F - FN'$. Consequently $v \sim 0$ on F mod $F''' + Y$, where $Y = Y' - FN_2$. But this implies that $u = v + w \sim 0$ on F mod $F'' + F''' + Y$, contrary to the fact that XC is of type A relative to (u, F, F'). The k-cycles (v, F, F'') and (w, F, F') are accordingly minimizing k-cycles. The fact that C_1 and C_2 are the critical k-sets determined by (v, F, F'') and (w, F, F') follows readily from Lemma 6.1. This proves Lemma 6.2.

7. Homotopic critical points. In Section 5 critical points were defined by means of homologies. In the present section it will be seen that these critical points are instances of a larger class of critical points that are determined by a special class of homotopic deformations, which we shall call F-deformations. In order to define F-deformations let us first recall the definition of homotopic deformations. Let P be a subset of S. To each point x in P and each value t on $0 \leqq t \leqq 1$ let there correspond a unique point $y(x, t)$ on S. The point function $y(x, t)$ $(0 \leqq t \leqq 1)$ will be said to define a *homotopic deformation* Δ *of* P *on* S if $y(x, 0) = x$ for each point x in P and the function $y(x, t)$ is continuous in its arguments. It will be called an *F-deformation* Δ *of* P if it has the further property that given a point x in P and a value t' on $0 \leqq t \leqq 1$, the points $y(x, t)$ $(t' \leqq t \leqq 1)$ belong to each set F in $\mathfrak{F}$ that contains the point $y(x, t')$. In the sequel it will be convenient to call the set formed by the points $y(x, 1)$, where x is in P, the *image* ΔP *of* P under the deformation Δ. Similarly, for a fixed value of t on $0 \leqq t \leqq 1$, the set of points of the form $y(x, t)$, where x is in P, will be denoted by the symbol $\Delta_t P$.

Consider now a set F in $\mathfrak{F}$. A point x' in F will be called a *homotopic critical point of* F if there does not exist, for any neighborhood N of x', an F-deformation Δ of FN such that for each value t on $0 < t \leqq 1$ the image $\Delta_t(FN)$ under Δ belongs to a primary subset F'' of F. The primary subset of F to be chosen depends on the value of t. This definition is analogous to one given by Morse (XVII, p. 419) and was suggested by it. The homotopic critical points of F are determined completely by F-deformations of the whole set F, as is readily seen by the use of the following:

LEMMA 7.1. *Let x' be a point on a set F in $\mathfrak{F}$ and suppose that x' is not a homotopic critical point of F. There exists a neighborhood N' of x' and an F-deformation Δ of F under which the points exterior to N' are held fast, while for each value t on $0 < t \leqq 1$ the image $\Delta_t P$ under Δ of a compact subset P of FN' is on a primary subset F_t of F.*

In order to prove this result let N be a neighborhood of the point x' for which there exists an F-deformation Δ of FN such that the set $\Delta_t(FN)$ is on a primary subset of F for each fixed value of t on $0 < t \leqq 1$. Such a neighborhood exists since x' is not a homotopic critical point of F. Let N' be an e-neighborhood of x' whose closure is in N and select a real single-valued continuous function $h(x)$ such that $h(x) = 0$ exterior to N' and $0 < h(x) \leqq 1$ on N'. For each value of t on $0 \leqq t \leqq 1$ set $z(x, t) = x$ exterior to N' and let $z(x, t) = y[x, h(x)t]$ on FN', where $y(x, t)$ is the point function defining

the F-deformation Δ. It is clear that the function $z(x, t)$ $(0 \leqq t \leqq 1)$ defines an F-deformation Δ' of F in which the points exterior to N' are held fast. Consider now a compact subset P of FN' and let m be the minimum of $h(x)$ on P. Since $h(x) > 0$ on P we have $m > 0$. Given a value t on $0 < t \leqq 1$ let F' be a primary subset of F such that $\Delta_{mt}(FN)$ is on F'. For each point x in P, the point $y(x, mt)$ must be on F' since $P \subset FN$. From the definition of F-deformations and the relation $h(x) \geqq m$ on P it follows that for each point x in P the point $z(x, t) = y[x, h(x)t]$ is also on F'. Consequently, the set $\Delta'_t P$ is on F' and the lemma is established.

COROLLARY. *The neighborhood N' of x' and the F-deformation of F described in Lemma 7.1 can be chosen so that for each value of t on $0 \leqq t \leqq 1$ the set $\Delta_t(FN')$ is on a prescribed neighborhood N of x'.*

It is clear from the proof of Lemma 7.1 that the neighborhood N' can be chosen so that its closure is in the prescribed neighborhood of x'. Moreover, the function $h(x)$ there used can be replaced by $rh(x)$, where r is a positive constant less than unity. When this has been done, it follows from continuity considerations that, for sufficiently small values of r, the F-deformation Δ will have the property described in the corollary.

THEOREM 7.1. *If C is the critical k-set determined by a non-null minimizing k-cycle (u, F, F'), then each point of C is a homotopic critical point of the set F belonging to (u, F, F').*

In order to prove this result suppose that there exists a point x' in C that is not a homotopic critical point relative to F. Let N' be a neighborhood of x' and let Δ be an F-deformation of F related to x' as described in Lemma 7.1. Let N'' be a neighborhood of x' whose closure is in N' and set $X = C - CN''$. In view of the definition of C, the set X cannot be of type B relative to (u, F, F'). Consequently there exists a subset X' of F of type A such that the set XX' is not of type A. There exists, accordingly, a neighborhood N^* of XX' and a primary subset F'' of F such that $u \sim 0$ on $F \bmod F'' + Y^*$, where $Y^* = F - FN^*$. By Lemma 2.2 we can suppose that the set U of vertices of u is on $F'' + Y^*$. Let N be a neighborhood of CX' whose closure is interior to the neighborhood $N^* + N'$ of CX'. We shall show that there exists a primary subset G of F such that the image ΔU of U under the F-deformation Δ is on $G + Y$, where $Y = F - FN$. To this end let $U'' = F''U$ and let U' be the closure of $U - U''$. Since Δ is an F deformation $\Delta U'' \subset F''$. The points of $Q = U' - U'N'$ are held fast under Δ and are at a positive distance from FN since Q is exterior to $N^* + N'$. There is, accord-

ingly, a neighborhood N_1 of Q such that $\Delta(U'N_1) \subset Y$. Finally the compact set $P = U' - U'N_1$ is on N' and, by Lemma 7.1, there exists a primary subset G of F such that $\Delta P \subset G$. We can suppose $G \supset F''$. Then $\Delta U = \Delta U'' + \Delta(U'N_1) + \Delta P$ is on $G + Y$, as was to be proved. Since the boundary u' is deformed on F' under Δ it follows from the theory of homotopic deformations that $u \sim 0$ on F mod $F' + G + Y$, contrary to the fact that C is a subset of F of type A, by Theorem 5.1. Our assumption that there exists a point x' in C that is not a homotopic critical point relative to F is accordingly false and Theorem 7.1 is established.

THEOREM 7.2. *There is no neighborhood N of a homotopic critical point x' of a set F in $\mathfrak{F}$ such that the set FN is on a subset of F. In particular a point x' in the critical k-set determined by a minimizing k-cycle (u, F, F') has this property.*

For suppose there is a neighborhood N of a homotopic critical point of F such that FN is on a proper subset F'' of F in $\mathfrak{F}$. Let Δ be the F-deformation of F in which all points of F are held fast. Then for each value t on $0 \leqq t \leqq 1$ the image $\Delta_t(FN)$ of FN is identical with FN and hence is on F''. This contradicts the definition of homotopic critical points of F. The first statement is accordingly established. The second follows from the fact that each point of the critical k-set C of a minimizing k-cycle (u, F, F') is a homotopic critical point of F.

Recalling the definition of admissible classes $\mathfrak{F}$ determined by a function $f(x)$, one obtains from the last theorem the following further result.

COROLLARY 1. *Let F be a set in the admissible class $\mathfrak{F}$ determined by a real single valued function $f(x)$, as described in* **3**. *Let b be the least upper bound of $f(x)$ on F. At each homotopic critical point x', the upper limit of $f(x)$ on F is equal to b. If $f(x)$ is continuous at x', then $f(x') = b$.*

COROLLARY 2. *Let $\mathfrak{F}$ be the admissible class determined by a real single valued lower semi-continuous function as described in* **3**. *Let C be the critical k-set determined by a minimizing k-set (u, F, F') and let b be the least upper bound of $f(x)$ on F. If there exists an F-deformation Δ of F such that $f(x)$ is continuous on ΔF, then $f(x) = b$ on C.*

8. Two lemmas. In the present section further properties of F-deformations will be developed. The results obtained will be useful in the next section. Let $P \subset Q$ be subsets of S and let Δ be an F-deformation of P defined by functions $y(x, t)$ $(0 \leqq t \leqq 1)$. The set P will be said to be *deformed on* Q under Δ

if the point $y(x, t)$ is on Q for each point x in P and each value t on $0 \leqq t \leqq 1$. The following lemma will be useful.

LEMMA 8.1. *Let F be a set in $\mathfrak{F}$ and let P be a compact subset of F containing no homotopic critical points of F. Given a neighborhood N of P there exists a neighborhood N^* of P and an F-deformation Δ of F having the following properties: The points of F exterior to N^* are held fast under Δ. The points of FN^* are deformed on FN under Δ. Finally, the image ΔP of P under Δ is on a primary subset F'' of F.*

In order to prove this result let N be a neighborhood of P. To each point x' of P let there correspond a neighborhood N' of x' and an F-deformation Δ related to x' and N' as described in Lemma 7.1 and its corollary. Since P is compact we can select a finite number of these neighborhoods $N_1, \cdots, N_q$ covering P. Let $\Delta_1, \cdots, \Delta_q$ be the corresponding F-deformations of F. Let Δ be the F-deformation of F obtained by applying successively the F-deformations $\Delta_1, \cdots, \Delta_q$. The F-deformation Δ and the neighborhood $N^* = N_1 + \cdots + N_q$ of P have the properties described in the lemma. The points exterior to N^* are held fast under each of the deformations $\Delta_1, \cdots, \Delta_q$ and hence also under Δ. Moreover, the set FN^* is deformed on FN under Δ since each of the deformations $\Delta_1, \cdots, \Delta_q$ has this property. It remains to show that ΔP is on a primary subset F'' of F. To do so we recall that under the F-deformation Δ, the points of $P - PN_1$ are held fast. There exists, accordingly, a neighborhood N_1'' of $P - PN_1$ such that the closure Q_2 of the image $\Delta_1(PN_1'')$ is on $N_2' = N_2 + \cdots + N_q$. By Lemma 7.1 with P replaced by $P = PN_1''$ there is a primary subset F_1 of F such that the set $Q_1 = \Delta_1(P - PN_1'')$ is on F_1. The set $P_1 = \Delta_1 P$ is accordingly the sum of two compact sets Q_1 and Q_2 such that Q_1 is on a primary subset F_1 on F and Q_2 is on N_2'. Repeating the argument made in the last four sentences with P replaced by Q_2, Δ_1 replaced by Δ_2, and N'_2 replaced by $N'_3 = N_3 + \cdots + N_q$, it is found that the set $\Delta_2 Q_2$ is the sum of two compact sets Q'_1 and Q'_2 such that Q'_1 is on a primary subset F'_1 of F and Q'_2 is on N'_3. The set $\Delta_2 Q_1$ is on F_1 since Δ_2 is an F-deformation. In the next repetition replace P by Q'_2, Δ_2 by Δ_3 and N'_2 by $N'_4 = N_4 + \cdots + N_q$. Proceeding in this manner it is found that the image ΔP of P under Δ is on a primary subset $F'' = F_1 + F'_1 + \cdots$ of F. This completes the proof of Lemma 7.2.

The proof of Theorem 9.2 is based on the following:

LEMMA 8.2. *Let F be a non-null set in $\mathfrak{F}$ and let H be an isolated set of homotopic critical points relative to F. Let N be a neighborhood of H*

containing in its closure no homotopic critical point of F that is not in H. Consider a k-cycle u on FN mod $F'N$, where F' is a primary subset of F. If $u \sim 0$ on F mod F', there is a primary subset F'' of F such that $u \sim 0$ on FN mod $(F' + F'')N$. The k-cycle (u, F, F') is minimizing if and only if there is no primary subset F'' of F such that $u \sim 0$ on FN mod $(F' + F'')N$.

In order to prove the first conclusion in the lemma let N', N'' be neighborhoods of H such that the closures of N', N'' are in N'', N respectively. Set $E = F(N'' - N')$. Given a k-cycle $u = (u_n)$ on FN mod $F'N$ such that $u \sim 0$ on F mod F', we shall show first that there is a k-cycle v on FN mod FN such that $u \sim v$ on FN mod $F'N$ and v is on $F'N + E$. To this end let $w = (w_n)$ be a $(k+1)$-chain such that $w' - u$ is on F'. We may replace the sequences (u_n), (w_n) by subsequences, if necessary, so that every simplex of u_n or of w_n having a point in N' is on N''. Let $\bar{w} = (\bar{w}_n)$ be the $(k+1)$-chain obtained from w by deleting all $(k+1)$-simplices having no point in common with N'. Then the k-chain $v = \bar{w}' - u$ is on $F'N + E$ and $u \sim v$ on FN mod $F'N$. Moreover, v is a k-cycle on FN mod $F'N$ since u has this property. Let V be the vertices of v and let P be the closure of the set $V - VF'$. If P is null, then $u \sim 0$ on FN mod $F'N$ and the first conclusion in the lemma holds with $F'' = 0$. If $P \neq 0$, then P is compact and contains no homotopic critical points since $P \subset N - N'$. Let Δ be an F-deformation related to P and N as described in the last lemma. Under Δ the points of F in a neighborhood N' of P are deformed on FN while the remaining points are held fast. Hence the set V is deformed on FN under Δ. Moreover, by Lemma 8.1, there is a primary subset F'' of F such that $\Delta P \subset F''$. The set $\Delta(V - P)$ is on F' since $V - P$ is on F'. Hence the set V is deformed under Δ on FN onto the set $F'N + F''N$. Consequently, $v \sim 0$ on FN mod $F'N + F''N$ and since $u \sim v$ on FN mod $F'N$ we have also $u \sim 0$ on FN mod $F'N + F''N$, as was to be proved.

The last statement in the lemma follows from the one just established. If the k-cycle (u, F, F') is not minimizing, there exists, by Lemma 3.1, a primary subset F'' of F such that $u \sim 0$ on F mod $F' + F''$. We have just seen that in this case F'' can be chosen so that $u \sim 0$ on FN mod $(F' + F'')N$. Conversely if there is a primary subset of F such that $u \sim 0$ on FN mod $(F' + F'')N$ then $u \sim 0$ on F mod $F' + F''$ and (u, F, F') cannot be minimizing. This completes the proof of Lemma 8.2.

9. F-homology classes and groups. Two minimizing k-cycles will be said to be *F-homologous* if they are homologous to a third minimizing k-cycle.

In view of Lemma 3.3 a minimizing k-cycle (u, F, F') with $F \neq 0$ is F-homologous to a k-cycle (v, G, G') if, and only if, the k-cycle $(u - v, F + G, F' + G')$ is not minimizing, that is, if, and only if, $F = G$ and there is a primary subset F'' of F such that $u \sim v$ on $F \bmod F''$. This last criterion suggests the term "F-homologous."

LEMMA 9.1. *The F-homology relation of minimizing k-cycles is reflexive, symmetric, and transitive.*

It is clear that the F-homology relationship is reflexive and symmetric. In order to prove that it is transitive let (u, F, F'), (v, G, G') and (v, G, G'), (w, H, H') be pairs of F-homologous minimizing k-cycles. Then $F = G = H$. If $F = 0$, then $(u, F, F') = (w, H, H') = (0, 0, 0)$. Suppose, therefore, that $F \neq 0$ and select a proper subset F'' of F in $\mathfrak{F}$ such that $u \sim v$ on $F \bmod F''$ and $v \sim w$ on $F \bmod F''$. Then $u \sim w$ on $F \bmod F''$ and (u, F, F') is F-homologous to (w, F, F'). This proves the lemma.

A class M_k composed of a minimizing k-cycle (u, F, F') and all minimizing k-cycles that are F-homologous to it will be called an *F-homology class.* The F-homology class composed of the minimizing k-cycle $(0, 0, 0)$ will be called the *zero F-homology class* $M_k = 0$. Let M_k be an F-homology class and let (u, F, F') be an k-cycle in M_k. The class containing $(-u, F, F')$ will be denoted by $-M_k$ and we set $M_k + (-M_k) = 0$. If $M'_k \neq -M_k$ is a second F-homology class and (v, G, G') is in M'_k, then the k-cycle $(u + v, F + G, F' + G')$ is a minimizing k-cycle belonging to a class which we define as the sum $M_k + M'_k$ of M_k and M'_k. It is easily seen that this definition of the sum $M_k + M'_k$ is independent of our choice of the k-cycles (u, F, F') and (v, G, G'). We have accordingly the following result.

THEOREM 9.1. *Under the above definition of addition the F-homology classes M_k define an abelian group $\mathfrak{M}_k$, called the k-th F-homology group of $\mathfrak{F}$.*

It is interesting to observe in passing that if F is composed only of the set $F = S$ and $F = 0$, the k-th F-homology group $\mathfrak{M}_k$ of $\mathfrak{F}$ is identical with the k-th homology group $\mathfrak{H}_k$ of S. An admissible class of this type is determined by the function $f(x) =$ constant on S. A relation between $\mathfrak{M}_k$ and $\mathfrak{H}_k$ when $\mathfrak{F}$ is arbitrary will be found in Theorem 10.2 below.

The k-th F-homology group $\mathfrak{M}_k$ is completely determined by the k-cycles on a neighborhood of the homotopic critical points, as is seen in the following:

THEOREM 9.2. *Let H be the set of all homotopic critical points on S determined by $\mathfrak{F}$. Let N be a neighborhood of H and S' the closure of N in S. The class $\mathfrak{F}'$ of all sets of the form FS', where F is in $\mathfrak{F}$, forms an admissible*

class relative to S'. Moreover, the k-th F-homology group $\mathfrak{M}'_k$ of $\mathfrak{F}'$ is isomorphic to the k-th F-homology group $\mathfrak{M}_k$ of $\mathfrak{F}$.

To prove this we note that every critical k-set C determined by a minimizing k-cycle (u, F, F') belongs to H, by virtue of Theorem 7. 1. Hence N is also a neighborhood of C. By Theorem 5. 2, with N replaced by S', there is a k-cycle v on $FS' \bmod F''S'$, such that $u \sim v$ on $F \bmod F''$. The system $(v, FS', F''S')$ is accordingly a minimizing k-cycle determined by the class $\mathfrak{F}'$. Let $(w, FS', F''S')$ be a second minimizing k-cycle determined by F' such that $u \sim w$ on $F \bmod F' + F'''$. Then $v \sim w$ on $F \bmod F'' + F'''$ and by Lemma 8. 2 with $u = v - w$, there is a primary subset G of F such that $v \sim w$ on $FN \bmod GN$. The k-cycles $(v, FS', F''S')$ and $(w, FS', F'''S')$ belong, therefore, to the same class M'_k in $\mathfrak{M}'_k$. Let the class M'_k correspond to the class M_k in $\mathfrak{M}_k$ containing (u, F, F'). Thus to each element M_k in $\mathfrak{M}_k$ there corresponds a unique element M'_k in $\mathfrak{M}'_k$. Consider next a k-cycle $(u, FS', F'S')$ in a class M'_k of $\mathfrak{M}'_k$. Then by Lemmas 8. 2 and 3. 1 the k-cycle (u, F, F') is also minimizing. It follows that each class M'_k in $\mathfrak{M}'_k$ is the image of one in $\mathfrak{M}_k$. The correspondence is obviously an isomorphism.

By virtue of Corollary 2 of Theorem 5. 1 each homology class M_k determines a unique critical k-set C, namely, the critical k-set determined by one of its members. As a further result we have

THEOREM 9. 3. *Let H be the set of all homotopic critical points on S and H_1 an isolated subset of H. Set $H_2 = H - H_1$. The F-homology classes whose critical k-sets are in H_1 form a subgroup $\mathfrak{M}_{k1}$ of $\mathfrak{M}_k$. Similarly the F-homology classes whose critical k-sets are in H_2 form a subgroup $\mathfrak{M}_{k2}$ and $\mathfrak{M}_k$ is the direct sum of $\mathfrak{M}_{k1}$ and $\mathfrak{M}_{k2}$. The groups $\mathfrak{M}_{k1}$ and $\mathfrak{M}_{k2}$ are completely determined by the minimizing k-cycles over neighborhoods N_1 and N_2 of H_1 and H_2 respectively.*

This result follows readily from the last theorem by taking N so small that S' is the sum of two sets S_1, S_2 containing H_1 and H_2 respectively.

10. Analogues of the critical point relations of Morse. The analogue of the critical point relations of Morse given in this section consists of relations between groups $\mathfrak{H}_k$, $\mathfrak{M}_k$, $\mathfrak{N}_k$ and $\mathfrak{P}_k$ $(k = 0, 1, \cdots)$, where $\mathfrak{H}_k$ is the k-th homology group of S, $\mathfrak{M}_k$ is the k-th F-homology group of $\mathfrak{F}$, and $\mathfrak{N}_k$, $\mathfrak{P}_k$ are the subgroups of $\mathfrak{M}_k$ described in Lemma 10. 1 below. More specifically, relations between the group $\mathfrak{H}_k$ and the difference groups $\mathfrak{N}_k - \mathfrak{P}_k$ and between the groups $\mathfrak{P}_k$ and $\mathfrak{M}_{k+1} - \mathfrak{N}_{k+1}$ are obtained. It will be seen in the next section that in case the group Λ used in defining chains is a field and $\mathfrak{M}_k$ is

of finite dimension, these relations imply that the groups $\mathfrak{H}_k$ and $\mathfrak{N}_k - \mathfrak{P}_k$ are of the same dimension. Similarly, the groups $\mathfrak{P}_k$ and $\mathfrak{M}_{k+1} - \mathfrak{N}_{k+1}$ are of the same dimension. The equality of the dimensions of these groups constitute the critical point relations of Morse, as will be seen in Theorem 11. 1 below.

LEMMA 10. 1. *The set of all F-homology classes M_k in $\mathfrak{M}_k$ containing a k-cycle of the form $(u, F, 0)$ forms a subgroup $\mathfrak{N}_k$ of $\mathfrak{M}_k$. Moreover $\mathfrak{M}_0 = \mathfrak{N}_0$. The set of all F-homology classes M_k in $\mathfrak{N}_k$ containing a k-cycle (u, F, F') such that $u \sim 0$ on S mod F'' for a suitably chosen proper subset F'' of F in $\mathfrak{F}$ forms, with $M_k = 0$, a subgroup $\mathfrak{P}_k$ of $\mathfrak{N}_k$.*

This result follows at once from the definition of the classes $\mathfrak{N}_k$ and $\mathfrak{P}_k$. It should be observed that every k-cycle (u, F, F') belonging to non-null class M_k in $\mathfrak{P}_k$ has the property that $u \sim 0$ on S mod F'' for a suitably chosen proper subset F'' of F in $\mathfrak{F}$. In fact every class M_k in $\mathfrak{M}_k$ containing a k-cycle with this property is necessarily in $\mathfrak{P}_k$.

The set F belonging to a minimizing k-cycle (u, F, F') will be called a minimizing k-set. Two k-cycles in the same F-homology class M_k determine the same minimizing k-set, which we shall denote by $F(M_k)$. We have the following:

THEOREM 10. 1. *Let F be a set in $\mathfrak{F}$. The non-zero F-homology classes in $\mathfrak{M}_k$, $\mathfrak{N}_k$, $\mathfrak{P}_k$ having $F(M_k) = F$ form, with $M_k = 0$, groups $\mathfrak{M}_k(F)$, $\mathfrak{N}_k(F)$, $\mathfrak{P}_k(F)$ respectively. The groups $\mathfrak{M}_k$, $\mathfrak{N}_k$, $\mathfrak{P}_k$, $\mathfrak{M}_k - \mathfrak{N}_k$, $\mathfrak{N}_k - \mathfrak{P}_k$ are direct sums over $\mathfrak{F}$ of the corresponding groups $\mathfrak{M}_k(F)$, $\mathfrak{N}_k(F)$, $\mathfrak{P}_k(F)$, $\mathfrak{M}_k(F) - \mathfrak{N}_k(F)$, $\mathfrak{N}_k(F) - \mathfrak{P}_k(F)$. In fact the direct sums need to be taken only over the class $\mathfrak{F}_k$ of all minimizing k-sets F.*

This result is an immediate consequence of the definition of the sum of F-homology classes given in the last section.

In view of the fact that the difference group $\mathfrak{N}_k - \mathfrak{P}_k$ is the direct sum over $\mathfrak{F}$ (or over $\mathfrak{F}_k$) of the difference groups $\mathfrak{N}_k(F) - \mathfrak{P}_k(F)$, one can obtain a satisfactory relationship between $\mathfrak{N}_k - \mathfrak{P}_k$ and $\mathfrak{H}_k$ by obtaining a relation between $\mathfrak{N}_k(F) - \mathfrak{P}_k(F)$ and $\mathfrak{H}_k$ for each set F in $\mathfrak{F}$ (or in $\mathfrak{F}_k$). This will be done in the following:

THEOREM 10. 2. *Let F be a set in $\mathfrak{F}$. Denote by $\mathfrak{H}_k(F)$ the set of all homology classes $\mathfrak{H}_k$ in the k-th homology group $\mathfrak{H}_k$ of S that contain a k-cycle u on F. Similarly denote by $\mathfrak{H}^*_k(F)$ the set of all homology classes in $\mathfrak{H}_k$ that contain a k-cycle on a proper subset of F in $\mathfrak{F}$. The sets $\mathfrak{H}_k(F)$ and $\mathfrak{H}^*_k(F)$ are subgroups of $\mathfrak{H}_k$ and the difference group $\mathfrak{H}_k(F) - \mathfrak{H}^*_k(F)$ is isomorphic with the difference group $\mathfrak{N}_k(F) - \mathfrak{P}_k(F)$ determined by the*

groups $\mathfrak{N}_k(F)$, $\mathfrak{P}_k(F)$ *described in Theorem* 10.1. *The difference group* $\mathfrak{H}_k(F) - \mathfrak{H}^*_k(F)$ *contains only the zero element of* $\mathfrak{H}_k$ *when* F *is not a minimizing* k*-set.*

It is clear that the sets $\mathfrak{H}_k(F)$, $\mathfrak{H}^*_k(F)$ form subgroups of $\mathfrak{H}_k$. In order to prove the remaining result consider first two homology classes H_k and H'_k belonging to the same non-zero coset of $\mathfrak{H}_k(F) - \mathfrak{H}^*_k(F)$. Let u, v be k-cycles on F belonging to $\mathfrak{H}_k$, $\mathfrak{H}'_k$, respectively. The k-cycles $(u, F, 0)$, $(v, F, 0)$ are minimizing. Suppose, for example, that the k-cycle $(u, F, 0)$ is not minimizing. Then it is homologous to a k-cycle $(w, F', 0)$ such that F' is a proper subset of F in $\mathfrak{F}$. The k-cycle w is in H_k and hence H_k belongs to $\mathfrak{H}^*_k(F)$, contrary to our choice of H_k as belonging to a non-zero coset of $\mathfrak{H}_k(F) - \mathfrak{H}^*_k(F)$. It follows that the k-cycle $(u, F, 0)$, and hence also $(v, F, 0)$, is minimizing. In view of the fact that the class $H_k - H'_k$ is in $\mathfrak{H}^*_k(F)$, there exists a proper subset F' of F in $\mathfrak{F}$ such that $u - v \sim 0$ on S mod F'. From this result we conclude that the F-homology classes M_k, M'_k containing the k-cycles $(u, F, 0)$, $(v, F, 0)$, respectively, belong to $\mathfrak{N}_k(F)$ and are such that $M_k - M'_k$ is in $\mathfrak{P}_k(F)$. The classes M_k, M'_k accordingly belong to the same coset of $\mathfrak{N}_k(F) - \mathfrak{P}_k(F)$. Moreover, this coset is not the zero coset $\mathfrak{P}_k(F)$. Otherwise there would exist a proper subset F' of F in $\mathfrak{F}$ such that $u \sim 0$ on S mod F' and the class H_k would be in $\mathfrak{H}^*_k(F)$, contrary to our assumption that H_k belongs to a non-zero coset of $\mathfrak{H}_k(F) - \mathfrak{H}^*_k(F)$. Since the classes H_k, H'_k need not be distinct, it has been shown that to each non-zero coset of $\mathfrak{H}_k(F) - \mathfrak{H}^*_k(F)$ there corresponds a unique non-zero coset $\mathfrak{N}_k(F) - \mathfrak{P}_k(F)$, related in the manner just described.

In order to show that this correspondence is one-to-one consider two F-homology classes M_k, M'_k belonging to the same non-zero coset of $\mathfrak{N}_k(F) - \mathfrak{P}_k(F)$ and let $(u, F, 0)$, $(v, F, 0)$ be k-cycles in M_k, M'_k, respectively. Since M_k is not in $\mathfrak{P}_k(F)$ the relation $u \nsim 0$ on S mod F' holds for every proper subset F' of F in $\mathfrak{F}$. Consequently the k-cycle u belongs to an homology class H_k in $\mathfrak{H}_k(F)$ which is not in $\mathfrak{H}^*_k(F)$. Similarly, the k-cycle v belongs to a second homology class H'_k of this type. From the fact that $M_k - M'_k$ is in $\mathfrak{P}_k(F)$, the relation $u - v \sim 0$ on S mod F' holds for a suitably chosen proper subset F' of F in $\mathfrak{F}$. It follows that $H_k - H'_k$ is in $\mathfrak{H}^*_k(F)$ and hence that H_k, H'_k belong to the same non-zero coset of $\mathfrak{H}_k(F) - \mathfrak{H}^*_k(F)$. The correspondence between the non-zero cosets of the difference groups $\mathfrak{H}_k(F) - \mathfrak{H}^*_k(F)$ and $\mathfrak{N}_k(F) - \mathfrak{P}_k(F)$ is therefore one-to-one. If the zero cosets are made to correspond, an isomorphism is obtained, as one readily verifies. This proves Theorem 10.2.

It should be observed that the group $\mathfrak{H}_k(F) - \mathfrak{H}^*_k(F)$ is essentially the

group of homology classes determined by k-cycles u on F which are non-bounding modulo every proper subset of F in $\mathfrak{F}$.

COROLLARY 1. *Suppose that the sets $F \neq S$ in $\mathfrak{F}$ are compact. Let $F \neq G$ be two subsets of $\mathfrak{F}$ such that $F \supset G$ and every minimizing k-set that is a proper subset of F is a subset of G. Then $\mathfrak{H}^*_k(F) = \mathfrak{H}_k(G)$ and $\mathfrak{H}_k(F) - \mathfrak{H}^*_k(F) = \mathfrak{H}_k(F) - \mathfrak{H}_k(G)$. Moreover $\mathfrak{H}_k = \mathfrak{H}_k(S)$.*

THEOREM 10.3. *Suppose that the sets $F \neq S$ in $\mathfrak{F}$ are compact and let G be a set in $\mathfrak{F}$. Let $\mathfrak{P}'_k(G)$ be the set of all F-homology classes M_k containing a k-cycle (u, F, F') such that $F \subset G$ and $u \sim 0$ on G mod F'' for a suitably chosen proper subset F'' of F in $\mathfrak{F}$. Let $\mathfrak{P}_k''(G)$ be the set of all F-homology classes M_k containing a k-cycle (u, F, F') for which there exist proper subsets F'', G'' of F, G in $\mathfrak{F}$ such that $u \sim 0$ on G'' mod F''. The sets $\mathfrak{P}_k'(G)$, $\mathfrak{P}_k''(G)$ are subgroups of $\mathfrak{P}_k$, $\mathfrak{P}_k'(G)$ respectively. Moreover, the difference group $\mathfrak{P}_k'(G) - \mathfrak{P}_k''(G)$ is isomorphic with the difference group $\mathfrak{M}_{k+1}(G) - \mathfrak{N}_{k+1}(G)$ determined by groups $\mathfrak{M}_{k+1}(G)$, $\mathfrak{N}_{k+1}(G)$ described in Theorem 10.1. These difference groups consist only of the zero coset when G is not a minimizing $(k+1)$-set.*

The first two statements in the theorem are immediate. In order to prove the last statement in the theorem let M_k, M_k' be F-homology classes belonging to the same non-zero coset of $\mathfrak{P}_k'(G) - \mathfrak{P}_k''(G)$. Let (u, F, F') and (v, H, H') be k-cycles in M_k, M_k', respectively, such that $u \sim 0$ on G mod F'', and $v \sim 0$ on G mod H'', where F'', H'' are proper subsets F, H in $\mathfrak{F}$. By part 1 of Lemma 2.3 there exist $(k+1)$-cycles w_1, w_2 on G mod F such that $u \sim w_1'$ on F mod F'', and $v \sim w_2'$ on H mod H''. We can suppose that $u = w_1'$, $v = w_2'$. The $(k+1)$-cycle (w_1, G, F) is minimizing, since otherwise there would exist a proper subset G'' of G in $\mathfrak{F}$ such that $w_1 \sim 0$ on G mod G''. This would imply that $u \sim 0$ on G'' and hence that M_k is in $\mathfrak{P}_k''(G)$, contrary to our choice of M_k. The F-homology class M_{k+1} containing (w_1, G, F) is in $\mathfrak{M}_{k+1}(G)$ but not in $\mathfrak{N}_{k+1}(G)$. If it were in $\mathfrak{N}_{k+1}(G)$ one could select a proper subset G'' of G and F'' of F in $\mathfrak{F}$, such that $w_1' = u \sim 0$ on G'' mod F'', contrary to our choice of u. Similarly the $(k+1)$-cycle (w_2, G, H) is a minimizing $(k+1)$-cycle belonging to an F-homology class M'_{k+1} in $\mathfrak{M}_{k+1}(G)$ which is not in $\mathfrak{N}_{k+1}(G)$. Since the class $M_k - M_k'$ is in $\mathfrak{P}_k''(G)$ there exist proper subsets F''' of $F + H$ and G'' of G in $\mathfrak{F}$ such that $u - v \sim 0$ on G'' mod F'''. From this property it follows readily that the classes M_{k+1}, M'_{k+1} belong to the same non-zero coset in $\mathfrak{M}_{k+1}(G) - \mathfrak{N}_{k+1}(G)$. Thus, to each non-zero coset of $\mathfrak{P}_k'(G) - \mathfrak{P}_k''(G)$ there corresponds a unique non-zero coset of $\mathfrak{M}_{k+1}(G) - \mathfrak{N}_{k+1}(G)$. In order to show that this correspondence is one to one, consider F-homology

classes M_{k+1}, M'_{k+1} belonging to a non-zero coset of $\mathfrak{M}_{k+1}(G) - \mathfrak{N}_{k+1}(G)$ and let (w_1, G, G'), (w_2, G, G'') be minimizing $(k+1)$-cycles in M_{k+1}, M'_{k+1} respectively. By Theorem 4.2, with $G = S$, one can assume that the k-cycles $(w_1', G', 0)$, $(w_2', G'', 0)$ are minimizing. These k-cycles then belong to F-homology classes M_k, M_k' in $\mathfrak{P}_k'(G)$. If M_k were in $\mathfrak{P}_k''(G)$, there would exist a proper subset G''' of G in $\mathfrak{F}$ such that $w_1' \sim 0$ on G''' mod G' and M_{k+1} would be in $\mathfrak{N}_{k+1}(G)$, which is not the case. Similarly, M_k' does not belong to $\mathfrak{P}_k''(G)$. On the other hand $M_k - M_k'$ is in $\mathfrak{P}_k''(G)$ since otherwise $M_{k+1} - M'_{k+1}$ could not be in $\mathfrak{N}_{k+1}(G)$, as one readily verifies. The correspondence between the groups $\mathfrak{M}_{k+1}(G) - \mathfrak{N}_{k+1}(G)$ and $\mathfrak{P}_k'(G) - \mathfrak{P}_k''(G)$ is accordingly one to one, when zero cosets are made to correspond. It is easily seen that the correspondence is an isomorphism. This completes the proof of Theorem 10.3.

COROLLARY. *Let $G \supset G'$ be sets in $\mathfrak{F}$ and suppose that every minimizing $(k+1)$-set contained in G is in G'. Then $\mathfrak{P}_k''(G) = \mathfrak{P}_k'(G')$ and the group $\mathfrak{P}_k'(G) - \mathfrak{P}_k''(G')$ is isomorphic with the group $\mathfrak{M}_{k+1}(G) - \mathfrak{N}_{k+1}(G)$, provided the sets $F \neq S$ in $\mathfrak{F}$ are compact. Moreover, $\mathfrak{P}_k'(S) = \mathfrak{P}_k$.*

11. The case when Λ is a field. In this section it is assumed that the group Λ is a field. The critical point relations of Morse can then be obtained in their original form. This is done in Theorem 11.1 below. First we make the following definitions and observations. If $u = \Sigma \lambda_i z_i$ is a vertex k-chain, we set $\lambda u = \Sigma(\lambda\lambda_i) z_i$. Given a Vietoris k-chain $u = (u_n)$ we define $\lambda u = (\lambda u_n)$. If (u, F, F') is a minimizing k-cycle, so also is $(\lambda u, F, F')$ when $\lambda \neq 0$, and they determine the same critical k-sets. Let M_k be an F-homology class containing (u, F, F'). By λM_k $(\lambda \neq 0)$ will be meant the class containing $(\lambda u, F, F')$. If $\lambda = 0$ we set $\lambda M_k = 0$. The relation $\lambda(M_k + M'_k) = \lambda M_k + \lambda M'_k$ holds for every pair of classes M_k and M'_k in $\mathfrak{M}_k$. The groups described in the last two sections, as well as the k-th homology group $\mathfrak{H}_k$, are distributive with respect to λ in this manner. The number of linearly independent elements (coefficients in Λ) in $\mathfrak{M}_k$ in a maximal set will be called the dimension m_k of $\mathfrak{M}_k$. The dimensions of the groups $\mathfrak{P}_k$ and $\mathfrak{N}_k - \mathfrak{P}_k$ will be denoted by p_k, r_k respectively. Suppose that m_k is finite when $k \leqq q$. Then for a fixed value of $k \leqq q$ for which $m_k > 0$ there exists a finite number of non-zero minimizing k-sets $F_1 \subset F_2 \cdot \cdot \cdot \subset F_n$. The groups $\mathfrak{M}_k(F_i)$, $\mathfrak{P}_k(F_i)$, $\mathfrak{N}_k(F_i) - \mathfrak{P}_k(F_i)$ have dimensions m_{ki}, p_{ki}, r_{ki}, respectively, and

$$m_k = m_{k1} + \cdot \cdot \cdot + m_{kn}, \quad p_k = p_{k1} + \cdot \cdot \cdot + p_{kn}, \quad r_k = r_{k1} + \cdot \cdot \cdot + r_{kn}.$$

By Theorem 10.2 and its corollary the dimension of the group $\mathfrak{H}_k(F_i) - \mathfrak{H}_k(F_{i-1})$ $(i > 0)$ is equal to r_{ki} and the dimension of $\mathfrak{H}_k(F_1)$ is r_{k1} provided the sets $F \neq S$ in $\mathfrak{F}$ are compact. Consequently, the dimension of the k-th ho-

mology group is equal to $r_{k1} + r_{k2} + \cdots + r_{kn} = r_k$. Similarly, by Theorem 10.3 and its corollary the dimension of the difference group $\mathfrak{P}'_{k-1}(F_i) - \mathfrak{P}'_{k-1}(F_{i-1})$ $(k > 0)$ is equal to the dimension $s_{ki} = m_{ki} - p_{ki} - r_{ki}$ of $\mathfrak{M}_k(F_i) - \mathfrak{N}_k(F_i)$ if $i > 1$ and the sets $F \neq S$ in $\mathfrak{F}$ are compact, as we shall assume. Moreover, the dimension of $\mathfrak{P}'_{k-1}(F_1)$ is equal to $s_{k1} = m_{k1} - p_{k1} - r_{k1}$. It follows that the dimension p_{k-1} of $\mathfrak{P}_{k-1}$ is equal to the dimension $s_{k1} + \cdots + s_{kn} = m_k - p_k - r_k$ of $\mathfrak{M}_k - \mathfrak{N}_k$. Hence $m_k = r_k + p_k + p_{k-1}$ $(0 < k \leqq q)$. Since $\mathfrak{M}_0 \equiv \mathfrak{N}_0$ we have $m_0 = r_0 + p_0$. This proves the following.

THEOREM 11.1. *Suppose that the sets $F \neq S$ in $\mathfrak{F}$ are compact. Let q be a positive integer and suppose that when $k \leqq q$ the dimension m_k of the k-th F-homology group $\mathfrak{M}_k$ is finite. Then the dimension p_k of the subgroup $\mathfrak{P}_k$ of $\mathfrak{M}_k$ $(k \leqq q)$ is finite and is equal to the dimension of the difference group $\mathfrak{M}_{k+1} - \mathfrak{N}_{k+1}$. Similarly the dimension r_k of the k-th homology group $\mathfrak{H}_k$ $(k \leqq q)$ is finite and is equal to the dimension of the difference group $\mathfrak{N}_k - \mathfrak{P}_k$. Moreover the relations*

$$(11.1) \qquad m_0 = r_0 + p_0, \quad m_k = r_k + p_k + p_{k-1} \qquad (k = 1, \cdots, q)$$

hold and the differences $h_k = m_k - r_k$ satisfy the inequalities

$$(11.2) \qquad h_k - h_{k-1} + \cdots + (-1)^k h_0 \geqq 0 \qquad (k = 0, 1, \cdots, q).$$

By the use of equation (11.1) it is found that the first member of (11.2) is identical with p_k and is accordingly non-negative, as was to be shown.

COROLLARY. *The number of isolated critical k-sets does not exceed m_k.*

In fact the number of isolated critical k-sets may equal m_k, as will be seen in **14** below.

Consider now an isolated set H of homotopic critical points and let $\mathfrak{M}^*_k$ be the class of all F-homology classes M_k whose critical k-sets are in H. As was seen in Theorem 9.3, the class $\mathfrak{M}^*_k$ is a subgroup of $\mathfrak{M}_k$ that is completely determined by the k-cycles on a neighborhood of H. Let m^*_k be the dimension of $\mathfrak{M}^*_k$. If $m^*_k > 0$ the set H will be said to be a homotopic critical k-set of index k of order m^*_k. The order m^*_k of H may be considered to be the number of critical k-sets in H. The set H may be of index k and index $j \neq k$ at the same time. By the use of Theorem 9.3 one obtains the following:

THEOREM 11.2. *Suppose that the dimension m_k of the k-th F-homology group is finite. The number of isolated homotopic critical sets of index k does not exceed m_k and the sum of their orders is equal to m_k.*

With a suitable interpretation this result is also true when m_k is infinite.

12. Extensions of admissible classes $\mathfrak{F}$. A theory analogous to that given above can be developed when the definition of admissible classes $\mathfrak{F}$ is modified in a number of ways. We shall give some examples and indicate briefly the modifications that have to be made. In the first place the assumption that the set $F = S$ is in $\mathfrak{F}$ is a trivial one. This assumption was used essentially to identify the dimensions of the groups $\mathfrak{N}_k - \mathfrak{P}_k$ and $\mathfrak{H}_k$ in Theorem 11.1. Similarly the assumption that the null set is in $\mathfrak{F}$ can be discarded if the set $F = 0$ is interpreted as the intersection of all sets in $\mathfrak{F}$. The assumption that the sets in $\mathfrak{F}$ are closed is not always essential. For example, for a continuous function $f(x)$ on S the set determined by inequalities of the form $f(x) \leqq c$ and $f(x) < c$ form a class for which the above theory is readily applicable. The theory so developed in analogous to one recently given by Seifert and Threlfall (XVIII).

A more significant modification is obtained if one replaces the hypothesis that given two sets in $\mathfrak{F}$ one is a subset of the other by the assumption that the sum of two sets in $\mathfrak{F}$ is in $\mathfrak{F}$. In this case we define a *primary subset* of F to be a subset F' of F in $\mathfrak{F}$ for which there exist no proper subset F'' of F in $\mathfrak{F}$ such that $F' + F'' = F$. The results described in **3** and **4** then hold with slight modification in their proofs. The results in **5** are also true although the sets X of type A relative to (u, F, F') can no longer be interpreted as those preserving the minimizing properties of (u, F, F') when the points exterior to a neighborhood of X are disregarded. However a set X of type A can be associated with the minimizing properties of a k-cycle (u, F, F') if one observes that the set Y appearing in the definition of such a set plays a role analogous to that of F'' in Lemma 3.1. The results described in **6** are not valid without further modification. However, those described in **7** and **8** hold true. In **9** the only modification that needs to be made is in the definition of the sums of F-homology classes. This modification is based on the fact that if (u, F, F') and (v, G, G') are minimizing k-cycles and H is the intersection of all sets in $\mathfrak{F}$ containing $F + G - FG$ as a subset then there is a minimizing k-cycle (w, H, H') with $H' = H(F' + G' + FG)$ such that $u + v \sim w$ on $F + G \bmod F' + G' + FG$. Moreover every k-cycle related to (u, F, F') and (v, F, F') in this manner is homologous to (w, H, H'). If $F = G$, then $(w, H, H') = (0, 0, 0)$. Thus if (u, F, F') is the F-homology class M_k and (v, F, F') is in the F-homology class M'_k, the class containing (w, H, H') will be defined as the sum $M_k + M'_k$ of the classes M_k and M'_k. With this modification the results described in **9** are found to be valid. Similarly the results described in the last two sections also hold, although the proofs need to be modified.

There are many examples of admissible classes of the type just described. For example, the class of all closed sets on S, or the class of all compact sets on S with the set $F = S$ adjoined are of this type. In this case the minimizing k-sets are minimum carriers of k-cycles . Similarly, given a homotopic deformation Δ of S the class of all closed sets F in S such that ΔF is deformed on F under Δ is a class of the type here considered. As a somewhat trivial but suggestive example of this case, let S be the lemniscate in the Euclidean two-dimensional space E defined by the equation $r^2 = \cos 2\theta$. We use the metric of E on S. Let O, A, B be respectively the points whose polar coordinates (r, θ) are $(0, 0)$, $(1, 0)$, $(1, \pi)$. Let D be the class of all homotopic deformations Δ of S on S which do not diminish the distance r of a point (r, θ) from the origin O. Let $\mathfrak{F}$ be the class of all closed subsets F of S such that $\Delta F \subset F$ for every deformation Δ in D. The class $\mathfrak{F}$ consists of the set S, the null set and the sums of the following sets: the points O, A, B; the closed subarcs of S having A or B as the initial point and O as the final end point; the closed subsets of S which have A or B as their initial point but do not pass through the point O. The points O, A, B are the only homotopic critical points. The minimizing k-sets are those for which $\Delta F = F$ for every deformation Δ in D. The points O, A, B and their sums are the minimizing 0-sets and critical 0-sets. The sums of the closed subarcs of S having A or B as their initial point and O as their final end-point are minimizing 1-sets and have the point O as their critical 1-set. Let Λ be the integers modulo 2. Then the connectivities r_k of S are given by the set $r_0 = 1$, $r_1 = 2$, $r_k = 0$ $(k \geqq 2)$ and the numbers m_k, p_k corresponding to those in **11** are $m_0 = 3$, $m_1 = 4$, $m_k = 0$ $(k \geqq 2)$, $p_0 = 2$, $p_1 = 2$, $p_k = 0$ $(k \geqq 2)$. The relations (11.1) and hence also (11.2) are accordingly satisfied.

13. Saddle points of functions. In this section we shall consider a real single valued function $f(x) = f(x_1, \cdots, x_n)$ defined over a neighborhood of the closure S of an open set in Euclidean $(x_1 \cdots x_n)$-space. It will be assumed that a neighborhood N of a point x on the boundary B of S can be represented in a one-to-one manner by continuous functions

$$x_i(\alpha, \beta) = x_i(\alpha_1, \cdots, \alpha_{n-1}, \beta) \qquad (i = 1, \cdots, n)$$

such that a point of N is in S if and only if $\beta \leqq 0$. The boundary points of S in N are given by the points $x_i(\alpha, 0)$. The function $g(\alpha) = f[x(\alpha, 0)]$ will be called the *boundary function*. The results described in this section hold also if S can be represented only locally by a Euclidean coordinate system.

Let x_0 be an interior point of S. The point x_0 will be called a *saddle*

point of $f(x)$ of index k if a neighborhood of x_0 can be represented in a one-to-one manner by continuous functions

$$x_i(y, z) = x_i(y_1, \cdots, y_k, z_1, \cdots, z_{n-k}) \qquad (i = 1, \cdots, n)$$

such that the point x_0 corresponds to the point $(y, z) = (0, 0)$ and such that the function $\phi(y, z) = f[x(y, z)]$ has the following two properties: (1) There exists a constant $r > 0$ such that for each point $(y, z) \neq (y, 0)$ satisfying the relations

$$(13.1) \qquad y_1^2 + \cdots + y_k^2 \leqq r^2, \qquad z_1^2 + \cdots + z_{n-k}^2 \leqq r^2$$

the function $\phi(y, hz)$ is an increasing function of h on the interval $0 \leqq h \leqq 1$; (2) For each $(hy, 0) \neq (0, 0)$ on the set (13.1) the function $\phi(hy, 0)$ is a decreasing function of h. From these two properties it is found that $\phi(0, z) > \phi(0, 0) > \phi(y, 0)$ for points $(0, z) \neq (0, 0)$ and $(y, 0) \neq (0, 0)$ on the set (13.1).

A boundary point of x_0 of S will be called a *saddle point of $f(x)$ of index k* if a neighborhood of x_0 can be represented by coordinates (y, z) as described above in such a way that the values (y, z) determine a point of S if and only if $z_{n-k} \leqq 0$ and such that properties (1) and (2) described above hold when $z_{n-k} \leqq 0$.

The admissible class $\mathfrak{F}$ of sets F here used is the one described in Example 1 of **6**, namely, the class of all subsets of S determined by an inequality of the form $f(x) \leqq c$, where c is a constant or $c = \pm \infty$. It will be assumed that the sets $F \neq S$ in $\mathfrak{F}$ are bounded. Since they are closed, they must also be compact and $f(x)$ has a finite lower bound on S. It will be assumed further that the coefficient group Λ is a field.

THEOREM 13.1. *Let x_0 be a saddle point of $f(x)$ of index k and let F be the set in $\mathfrak{F}$ determined by the inequality $f(x) \leqq f(x_0)$. The point x_0 is a homotopic critical point relative to F and there is a neighborhood of x_0 containing no other homotopic critical point. In fact the point x_0 forms a homotopic critical set of index k and order one and is not of index $j \neq k$. The index of a saddle point is uniquely defined.*

In order to prove this result we shall restrict ourselves to a saddle point x_0 that is interior to S. The saddle points on the boundary can be treated similarly by restricting the coordinate z_{n-k} to be of constant sign. Let a neighborhood of the point x_0 under consideration be represented by the coordinates (y, z) described above. Denote by c the minimum of the values $\phi(y, z)$ on the set

$$y_\alpha y_\alpha = 0, \qquad z_\sigma z_\sigma = r^2 \qquad (\alpha = 1, \cdots, k;\ \sigma = 1, \cdots, n-k).$$

Here and elsewhere a repeated subscript will be used to denote summation with respect to that subscript. Choose a constant h such that $\phi(0,0) < h < c$ and let H be the set in $\mathfrak{F}$ determined by the inequality $f(x) \leqq h$. Select a positive constant $s < r$ such that the relation $\phi(y,z) > h$ holds on the set

$$y_\alpha y_\alpha \leqq s^2, \qquad z_\sigma z_\sigma = r^2,$$

and let N be the neighborhood of x_0 determined by the inequalities

$$y_\alpha y_\alpha < s^2, \qquad z_\sigma z_\sigma < r^2.$$

We shall show now that there exists an F-deformation Δ of H which deforms the set $H - HN$ on $H - HN$ and deforms the set HN on HN into the set determined by the relations

$$(13.2) \qquad y_\alpha y_\alpha < s^2, \qquad z_\sigma = 0 \qquad (\alpha = 1, \cdots, k;\ \sigma = 1, \cdots, n-k).$$

To do so let $h(y)$ be a real valued continuous function of $y = (y_1, \cdots, y_k)$ such that $h(y) = 1$ when $y_\alpha y_\alpha \leqq s^2$, $0 \leqq h(y) \leqq 1$ when $s^2 \leqq y_\alpha y_\alpha \leqq r^2$ and $h(y) = 0$ elsewhere. Let Δ be the F-deformation of H defined by the following two properties: The points determined by the values of (y,z) on (13.1) are replaced by the points determined by the values $(\bar{y}, \bar{z})$ according to the law

$$\bar{y}_\alpha = y_\alpha, \qquad \bar{z}_\sigma = (1-t)h(y)z_\sigma \qquad (0 \leqq t \leqq 1);$$

all other points of H are held fast. On N the values (y,z) are replaced by the values $(y, (1-t)z)$ which are also on N. Since the function $f[y, (1-t)z]$ is a decreasing function of t for fixed values of $(y,z) \neq (y,0)$, it follows that there is no homotopic critical point on HN except possibly on the set (13.2). We shall see presently that the point $(y,z) = (0,0)$ is the only homotopic critical point on N. It should be observed that if $k = 0$ the points of HN are deformed into a single point under Δ. The conclusions of the theorem are therefore immediate when $k = 0$. In the sequel we shall suppose that $k > 0$.

Let P be the set obtained by deleting the point x_0 from the set F described in the theorem. The points of the image ΔP of P under the F-deformation Δ which lie in N are determined by the relations

$$(13.3) \qquad 0 < y_\alpha y_\alpha < s^2, \qquad z_\sigma = 0.$$

Let Δ' be the F-deformation of ΔP for which the points of ΔP that are not in N are held fast and such that a point (y,z) in N and hence on the set (13.5) is replaced by the point $(\bar{y}, \bar{z})$ determined by the relations

$$\bar{y}_\alpha = y_\alpha(1 - t + ts/(y_\beta y_\beta)^{1/2}), \qquad \bar{z}_\sigma = 0 \qquad (0 \leqq t \leqq 1).$$

The deformation Δ of P followed by the deformation Δ' of ΔP defines an F-deformation Δ'' of P which carries the points of PN into the $(k-1)$-sphere

$$(13.4) \qquad y_\alpha y_\alpha = s^2, \qquad z_\sigma = 0 \qquad (\alpha = 1, \cdots, k;\ \sigma = 1, \cdots, n-k).$$

It follows that there is no homotopic critical point on the set PN. The point x_0 is accordingly the only possible critical point in HN.

We shall now prove that the point x_0 forms a critical k-set and hence also a homotopic critical set of index k. To do so consider a j-cycle (u, F, F'), where F is the set described in the theorem. Since the value of $\phi(y, z)$ is less than $\phi(0, 0)$ on the set (13.4) we can suppose that the set (13.4) is on the set F' without disturbing the minimizing properties of (u, F, F'). Under the F-deformation Δ described above the j-chain u is deformed into a j-chain v on ΔF such that $u \sim v$ on $F \bmod F'$. We can accordingly suppose that u is on the set ΔF since we can replace u by v. Let $Y = F - FH$. In view of the definition of critical j-sets we are interested only in determining when the relation $u \nsim 0$ on $F \bmod F' + F'' + Y$ holds for every proper subset F'' of F in $\mathfrak{F}$. We can accordingly suppose that u is on the closure of the set $\Delta(FN)$, that is, on the set

$$(13.5) \qquad y_\alpha y_\alpha \leqq s^2, \qquad z_\sigma = 0.$$

Under the deformation Δ'' of F' determined by the deformation Δ'' of P the boundary u' of u is deformed into a $(k-1)$-chain on the set (13.4). We can suppose, therefore, that u' is on the set (13.4). If $u' \sim 0$ on the set (13.4) we can suppose, by Lemma 2.3, that $u' = 0$ and that u is a j-cycle on the set (13.5). Since the j-th connectivity $(j > 0)$ of the set (13.5) is zero, we have $u \sim 0$ on this set and hence $(u, F, F') \sim 0$ also, in this case. It follows that the point x_0 cannot form a critical j-set $(j \neq k)$. In order to show that the point x_0 forms a critical k-set we observe that there is a $(k-1)$-cycle v' on the set (13.4) which is non-bounding on this set. This $(k-1)$-cycle bounds a k-chain v which is a k-cycle on $F \bmod F'$. In fact if F'' is any proper subset of F in $\mathfrak{F}$ containing F', then $v \nsim 0$ on $F \bmod F''$. For otherwise we would have $v \sim 0$ on $\Delta F \bmod \Delta F''$ and hence $v \sim 0$ on $\Delta(FN) \bmod \Delta(F''N)$. Under the deformation described above the set $\Delta(F''N)$ is deformed into the set (13.4). It would follow that we would have $v' \sim 0$ on the set (13.4), contrary to our choice of v'. The k-cycle (v, F, F'') is, therefore, a minimizing k-cycle and x_0 is the critical k-set determined by it. Moreover there is an element λ in the field Λ such that $u' \sim \lambda v'$ on the set (13.4). We have accordingly $(u, F, F') \sim (\lambda v, F, F')$. The count of the point x_0 as a critical k-set is, therefore, unity. This completes the proof of Theorem 13.1.

THEOREM 13.2. *Suppose that the homotopic critical points of $f(x)$ are all saddle points and let m_k be the number of saddle points of $f(x)$ of index k on S. If the numbers m_k $(k = 0, 1, \cdots, n)$ are finite, then the connectivities $r_0, \cdots, r_n$ of S are finite and the numbers $h_k = m_k - r_k$ satisfy the relations* (2.3) *and*

$$h_n - h_{n-1} + h_{n-2} - \cdots + (-1)^n h_0 = 0. \tag{13.6}$$

If S is bounded and closed and if no point of S is a limit point of homotopic critical points, these numbers are finite.

14. Analytical critical points of functions. In the present section it will be assumed that the function $f(x)$ described in the last section has continuous second partial derivatives on a neighborhood of the closed S. It will be assumed further that the functions $x_i(\alpha_1, \cdots, \alpha_{n-1})$ $(i = 1, \cdots, n)$ locally defining the boundary of S have continuous second derivatives and that the matrix of their first derivatives has rank $n - 1$. The boundary function $g(\alpha) = f[x(\alpha)]$ has continuous second derivatives.

A point x_0 at which the first partial derivatives of $f(x)$ are all zero will be called an analytic critical point of $f(x)$. Similarly the analytic critical points of the boundary function $g(\alpha)$ are those at which the first partial derivatives vanish.

THEOREM 14.1. *Let x_0 be a homotopic critical point of $f(x)$ on S. If x_0 is interior to S, it is an analytic critical point of $f(x)$. If x_0 is on the boundary B of S it is an analytic critical point of the boundary function g and the outer normal derivative f_N of $f(x)$ at $x = x_0$ satisfies the inequality $f_N \leqq 0$, the equality holding only in case the point x_0 is also an analytic critical point of $f(x)$.*

In order to prove the first statement suppose that one of the derivatives, say f_{x_1}, is not zero at $x = x_0$. Replace x_1 by $-x_1$ if necessary, so that $f_{x_1} > 0$ on a neighborhood N of x_0. Let Δ be the F-deformation of N defined by the system

$$y_1 = x_1 - bt, \qquad y_j = x_j \qquad (j = 2, \cdots, n;\ 0 \leqq t \leqq 1)$$

where b is a small positive constant. The function $f(y)$ thus obtained is a decreasing function of t for each point x in N. This contradicts our definition of homotopic critical points. The point x is, accordingly, an analytic critical point of $f(x)$. The second statement in the theorem can be proved similarly.

An analytic critical point x_0 of $f(x)$ will be said to be *non-degenerate* if

the determinant $|f_{x_ix_j}|$ $(i, j = 1, \cdots, n)$ is different from zero at $x = x_0$. By the *index* of a non-degenerate analytic critical point x_0 will be meant the negative index of the quadratic form $f_{x_ix_j}\xi_i\xi_j$ $(i, j = 1, \cdots, n)$ at $x = x_0$.

THEOREM 14.2. *A non-degenerate analytic critical point x_0 of $f(x)$ of index k, that is interior to S, is a saddle point of index k.*

For suppose that the point x_0 is the origin and that the coordinate system has been chosen so that at $x = x_0$ we have

$$f_{x_ix_j}\xi_i\xi_j = -\xi_1^2 - \cdots - \xi_k^2 + \xi_{k+1}^2 + \cdots + \xi_n^2. \tag{14.1}$$

The $n - k$ independent equations

$$f_{x_j}(x_1, \cdots, x_k, x_{k+1}, \cdots, x_n) = 0 \qquad (j = k+1, \cdots, n) \tag{14.2}$$

hold at $x = 0$ and hence have unique solutions $x_j(x_1, \cdots, x_k)$ $(j = k+1, \cdots, n)$ which are continuous and have continuous derivatives near the values $(x_1, \cdots, x_k) = (0, \cdots, 0)$. The function $\phi(y, z)$ obtained from $f(x)$ by setting

$$x_\alpha = y_\alpha, x_{k+\sigma} = x_{k+\sigma} = x_{k+\sigma}(y_1, \cdots, y_k) + z_\sigma \quad (\alpha = 1, \cdots, k; \sigma = 1, \cdots, n-k)$$

can be shown to have the properties described in the definition of a saddle point of index k. In the first place we have

$$\phi_{z\sigma}(y, 0) = f_{x_{k+\sigma}}[y_1, \cdots, y_k, x_{k+1}(y), \cdots, x_n(y)] = 0. \tag{14.3}$$

By the use of Taylor's theorem it is found that

$$\phi(y, hz) = \phi(y, 0) + h^2 \int_0^1 (1-t)\phi_{z\sigma z\tau}(y, tz) z_\sigma z_\tau dt. \tag{14.4}$$

The integrand is positive at a point $(y, z) \neq (y, 0)$ on a set of the form

$$y_\alpha y_\alpha \leqq r^2, \qquad z_\sigma z_\sigma \leqq r^2 \qquad (\alpha = 1, \cdots, k;\ \sigma = 1, \cdots, n-k), \tag{14.5}$$

by virtue of the formula $\phi_{z\sigma z\tau} = f_{x_{k+\sigma}x_{k+\tau}}$ and the relation (14.5) with $\xi_1 = \cdots = \xi_k = 0$. The function $\phi(y, hz)$ is accordingly an increasing function of h on the set (14.5). On the other hand the function $\phi(y, 0)$ has the continuous first and second derivatives

$$\phi_{y_\alpha} = f_{x_\alpha} + f_{x_j}x_{jy_\alpha} = f_{x_\alpha}[y_1, \cdots, y_k, x_{k+1}(y), \cdots, x_n(y)],$$
$$\phi_{y_\alpha y_\beta} = f_{x_\alpha x_\beta} + f_{x_\alpha x_j}x_{jy_\beta} \quad (\alpha, \beta = 1, \cdots, k;\ j = k+1, \cdots, n).$$

Differentiating the members of the equation (14.3) with respect to y_α it is found that

$$f_{x_jx_\alpha} + f_{x_jx_l}x_{ly_\alpha} = 0 \qquad (j, l = k+1, \cdots, n).$$

We have accordingly

$$\phi_{y_\alpha y_\beta} = f_{x_\alpha x_\beta} - f_{x_j x_i} x_{iy_\alpha} x_{jy_\beta}.$$

It follows from this relation and formula (14.1) that the quadratic form $\phi_{y_\alpha y_\beta}\xi_\alpha\xi_\beta$ $(\alpha, \beta = 1, \cdots, k)$ is negative definite at $(y, z) = (0, 0)$. From the formula

$$\phi(hy, 0) = \phi(0, 0) + h^2 \int_0^1 (1-t)\phi_{y_\alpha y_\beta}(ty, 0) y_\alpha y_\beta \, dt$$

we conclude that the function $\phi(hy, 0)$ is a decreasing function of h for each point $(hy, 0) \neq (0, 0)$ on the set (14.5), provided that r is taken sufficiently small. The point x_0 is therefore a saddle point of $f(x)$ of index k, as was to be proved.

By a similar argument one can establish the following:

THEOREM 14.3. *A non-degenerate analytic critical point of index k of the boundary function $g(\alpha)$, at which the outer normal derivative f_N of $f(x)$ is negative is a saddle point of $f(x)$ of index k.*

Combining Theorems 14.2, 14.3 and 13.2 one obtains the following result, first proved by Morse (III) in a simpler form and later extended to the present theorem by Morse and Van Schaack (IX).

THEOREM 14.4. *Suppose that the analytic critical points of $f(x)$ on S are non-degenerate and interior to S. Suppose further that the analytic critical points of the boundary function $g(\alpha)$, at which the outer normal derivative f_N of $f(x)$ is negative, are also non-degenerate. Let m_k be the number of analytic critical points of $f(x)$ of index k on S plus the number of analytic critical points of $g(\alpha)$ of index k at which $f_N < 0$. If the numbers $m_0, \cdots, m_n$ are finite, so also are the connectivities $r_0, \cdots, r_n$ of S and the numbers $h_k = m_k - r_k$ satisfy the relations (2.3) and (12.6). If the set S is bounded, these numbers are necessarily finite.*

II. APPLICATIONS TO THE CALCULUS OF VARIATIONS.

In the next two sections an application of the above theory will be made to a simple problem in the Calculus of variations. The present treatment is a brief one but is sufficient to bring out some of the essential processes and ideas that are useful in a more extensive treatment. For example, in order to show that every k-cycle is homologous to a minimizing k-cycle it is sufficient

to show that the hypotheses of Theorem 4. 1 are satisfied, namely, that the sets $f \leqq c$ are compact. The existence of absolute minima is usually established by imposing conditions on the function under consideration which will imply that the sets $f \leqq c$ are compact. Thus it is seen that the results obtained by the various writers on absolute minima can be applied here so as to obtain criteria for the existence of minimizing k-cycles and minimizing k-sets. Moreover in **16** below it is shown that in the positive regular case critical arcs are extremal arcs. This is done by the use of the theory of short extremals. This theory played an important role in the earlier existence theorems for absolute minima. It was used by Birkhoff in order to establish the existence of extremals of index unity. It plays a basic role in the earlier papers of Morse and in the abstract variational theory recently developed by Morse.

15. The positive quasi-regular case. Let $\mathfrak{X}$ be a bounded closed set of points $x = (x_1, \cdots, x_n)$ in an Euclidean n-space. Denote by $\phi(x, p) = \phi(x_1, \cdots, x_n, p_1, \cdots, p_n)$ a real single valued function defined for all elements $(x, p) \neq (x, 0)$ with x in a neighborhood $\mathfrak{N}$ of $\mathfrak{X}$. It will be assumed that for these values of (x, p) the function $\phi(x, p)$ has continuous derivatives of the first three orders and is positively homogeneous of degree one in p. It will be assumed further that for each set (x, p, q) with x in $\mathfrak{X}$ and $p \neq 0$, $q \neq 0$ one has

$$(15.1) \qquad \phi(x, p) > 0, \qquad E(x, p, q) = \phi(x, q) - q_i \phi_{p_i}(x, p) \geqq 0,$$

where $(i = 1, \cdots, n)$. These relations are the so-called conditions of *positive definiteness* and *positive quasi-regularity* [2] for the integral

$$f(X) = \int_{t_1}^{t_2} \phi[x(t), x'(t)]dt.$$

Here X denotes an absolutely continuous curve

$$(15.2) \qquad x_i(t) \qquad (t_1 \leqq t \leqq t_2;\ i = 1, \cdots, n)$$

on $\mathfrak{N}$. In this section and the following one the argument of the function f and hence the points of the space S of the preceding pages will be denoted by capital letters instead of small letters.

Let S^* be the class of all directed rectifiable curves X in $\mathfrak{X}$. Curves consisting of a single point are admitted. Two curves in S^* will be considered identical if they consist of the same points in the same order. We introduce a Fréchet distance on S^* in the usual manner (VII, p. 202). A metric space

[2] See for example, McShane (VII, p. 204). Here references are found to the important works of Tonelli, Graves, and others.

S^* is thus obtained. Denote by S a subspace of S^* that is closed relative to S^* and let $\mathfrak{F}$ be the admissible class of sets F on S determined by inequalities of the form $f(X) \leqq c$, where c is a real constant or $+\infty$.

THEOREM 15.1. *The sets $F \neq S$ in $\mathfrak{F}$ are compact.*

It is well known that the conditions (15.1) imply that given a finite constant c there is a constant N such that if X is a curve in S having $f(X) \leqq c$, then $L(X) \leqq N$, where $L(X)$ is the length of X. Moreover the function $f(X)$ is lower semi-continuous on the subset of curves X in S defined by the inequality $L(X) \leqq N$. These results follow, for example, from the proofs of Theorem 1 and Lemma 3 of McShane (VII, p. 209). Consider now a sequence of curves $\{X_r\}$ in S having $f(X_r) \leqq c$. In view of the relations $L(X_r) \leqq N$, the sequence has a rectifiable accumulation curve X_0, by Hilbert's theorem (cf. VII, p. 210). The curve X_0 is in S since S is closed in S^*. Moreover $f(X_0) \leqq \lim f(X_r) \leqq c$, by virtue of the lower semi-continuity of $f(X)$ on the set of curves X having $L(X) \leqq N$. This proves the theorem.

In view of the last theorem it is seen by Theorems 4.1 and 5.1 that critical k-sets exist whenever modular k-cycles exist that are not homologous to zero. Further properties of critical k-sets are yet to be determined. For example it is not known whether the above hypotheses are sufficient to show that on the critical k-set C corresponding to a non-null minimizing k-set F in $\mathfrak{F}$, the relation $f(X) = c$ holds, where c is the least upper bound of $f(X)$ on F. This result is true in the positive regular case as will be seen in the next section.

16. The positive regular case. In the present section we shall make the additional assumption that at each element $(x, p) \neq (x, 0)$ with x in $\mathfrak{R}$ the inequality

$$\phi_{p_i p_k} h_i h_k > 0 \qquad (i, k = 1, \cdots, n) \tag{16.1}$$

holds for every set $(h) \neq (bp)$. This is the so-called *positive regular case.* In this case extremals exist. We shall show that every homotopic critical curve, and hence that every critical curve of index k, is an extremal arc.

An arc X on $\mathfrak{R}$ that can be represented by functions (15.2) having continuous first and second derivatives with $x'_i x'_i \neq 0$ will be called an *extremal arc* if it satisfies the Euler equations

$$\phi_{x_i} - (d/dt)\phi_{p_i} = 0.$$

An extremal X will be called a minimizing extremal if it affords a minimum

to $f(X)$ in the class of all rectifiable curves on $\mathfrak{R}$ joining its end points and is similarly directed. It is well known that the hypotheses made above insure the existence of a constant $e > 0$ such that every pair of points in $\mathfrak{X}$ within an e-distance of each other can be joined in either direction by a unique minimizing extremal on $\mathfrak{R}$. These extremals vary continuously with their end points and the integral $f(X)$ is continuous on this class of extremal arcs. We suppose that a constant e has been so chosen.

In the present section it will be assumed that the subclass S of S^* described in **15** has the following additional property: *Let Y be a subarc of a curve X in S such that the diameter of Y does not exceed the constant e chosen above. The space S contains the arc obtained from X by replacing the subarc Y of X by the unique minimizing extremal joining its end points and similarly directed.* This assumption is essentially an assumption of extremal convexity of our space S.

In setting up the F-deformations described below we shall use a special parametrization of curves due to Whitney and Morse (XV). They have shown that a curve X in S can be defined by points

$$x_i(t, X) \qquad (0 \leqq t \leqq 1;\ i = 1, \cdots, n), \tag{16.2}$$

where the functions $x_i(t, X)$ are continuous in t and X for all t on $0 \leqq t \leqq 1$ and X in S. Moreover for each fixed curve X, not a point curve, the functions (15.2) are not all constants on any subinterval of $0 \leqq t \leqq 1$.

LEMMA 16.1. *Let P be a compact subset of S and suppose that the arcs of S have been parameterized in the manner described above. For every constant $\epsilon > 0$ there is a constant $\delta > 0$ such that if $t' \leqq t \leqq t''$ is a subinterval of $0 \leqq t \leqq 1$ of length at most δ, the diameter of the subarc of a curve X in P determined by the subinterval $t' \leqq t \leqq t''$ cannot exceed ϵ.*

This lemma follows readily from the fact that the functions $x_i(t, X)$ in (16.2) are continuous on the compact set of elements (t, X) with t on $0 \leqq t \leqq 1$ and X on P and hence are uniformly continuous on this set.

The results describe below are obtained by the use of two F-deformations $\Delta(t', t'')$ and Δ of a compact set P on S. To define the first of these, consider a compact set P of S and suppose that the arcs of P have the special parametrization described above. Let e be the constant chosen at the beginning of this section and choose δ related to $\epsilon = e$ as described in Lemma 15.1. Let $t' \leqq t \leqq t''$ be a subinterval of $0 \leqq t \leqq 1$ of length at most δ. Denote by $Y(\tau, X)$ $(t' \leqq \tau \leqq t'')$ the curve obtained from a curve X in P by replacing the subarc of X determined by $t' \leqq t \leqq \tau$ by the minimizing extremal joining

its end points and similarly directed. We have $Y(t', X) = X$. Since these short minimizing extremals vary continuously with their end points, the function $Y(\tau, X)$ is continuous on S for all values (τ, X) with τ on $t' \leqq \tau \leqq t''$ and X on P. The function $Y(\tau, X)$ $(t' \leqq \tau \leqq t'')$ therefore defines a homotopic deformation $\Delta(t', t'')$ of P on S.

LEMMA 16.2. *The deformation $\Delta(t', t'')$ of a compact set P on S is an F-deformation of P on S. Under this deformation the subarc of a curve X in P determined by the interval $t' \leqq t \leqq t''$ is replaced by a minimizing extremal arc while the remaining points of X are held fast.*

By the use of the deformation $\Delta(t', t'')$ one can define an F-deformation Δ of a compact subset P of S as folliows: Let q be an integer such that $q\delta > 1$, where δ is the constant used in the definition of $\Delta(t', t'')$. Divide the interval $0 \leqq t \leqq 1$ into q equal subintervals by values $t_0 = 0 < t_1 < t_2 < \cdots < t_q = 1$. We then have $0 < t_j - t_{j-1} < \delta$ $(j = 1, \cdots, q)$. The deformation obtained by applying the F-deformations $\Delta(t_{j-1}, t_j)$ $(j = 1, \cdots, q)$ simultaneously to the arcs of P will be called the deformation Δ of P on S. Under this deformation the subarc of X determined by $t_{j-1}t_j$ is replaced by an extremal segment. The curve X is therefore deformed into a broken extremal with at most $q - 1$ corners. On this class of broken extremals the integral $f(X)$ is continuous. Hence we have:

LEMMA 16.3. *Under the deformation Δ the compact set P is deformed into a compact set ΔP on which the integral $f(X)$ is continuous.*

The least upper bound b of $f(X)$ on a minimizing k-set F will be called a *critical value of index k corresponding to F*. The set F is accordingly determined by the inequality $f(X) \leqq b$. As a first result we have:

THEOREM 16.1. *The set $F = S$ is not a minimizing k-set, that is, the critical values of index k which exist are finite.*

For consider a k-cycle (u, F, F') with $F = S$ and let (u_n) be a sequence of vertex k-chains defining u. Let w be a $(k+1)$-chain defined by a sequence (w_n) of vertex $(k+1)$-chains such that $u_n - u_{n+1} - w'_n$ is on F'. The vertices of the chains in these sequences together with their limit points form a compact set P. Under the F-deformation Δ of P described above, the set P is deformed onto a set ΔP on which f is continuous and hence bounded. There is, accordingly, a proper subset G of F in $\mathfrak{F}$ such that $\Delta P \subset G$. Under the deformation Δ the k-chain u is deformed into a k-chain v on G. It follows that $u \sim 0$ on $F = S \bmod G$. The k-cycle (u, F, F') is, therefore, not a minimizing k-cycle. This proves the theorem.

THEOREM 16.2. *Let F be a non-null minimizing k-set and let b be the least upper bound of $f(X)$ on F, that is, the critical value of index k corresponding to F. On the critical k-set C corresponding to F, $f(X) = b$.*

This result follows at once from Corollary 2 of Theorem 7.2 and the properties of the F-deformation Δ of F described in Lemma 16.3.

THEOREM 16.3. *Every homotopic critical curve corresponding to a set $F \neq S$ in $\mathfrak{F}$ and hence every curve in a critical k-set is an extremal arc.*

For let X be a critical curve corresponding to a set $F \neq S$ in $\mathfrak{F}$. The curve X is therefore a fixed curve under the F-deformation $\Delta(t', t'')$ of $P = F$ described above. The subarc of X, determined by an arbitrary interval $t' \leqq t \leqq t''$ of $0 \leqq t \leqq 1$ of length at most δ, is accordingly an extremal arc. It follows that X must also be an extremal, as was to be proved.

THEOREM 16.4. *Let S be the class of all rectifiable curves X on $\mathfrak{X}$ joining two fixed points and let X_0 be an extremal in S on the interior of $\mathfrak{X}$. Suppose that the end points of X_0 are not conjugate and denote by k the number of points on X conjugate to its initial end point, each conjugate point being counted a number of times equal to its order. The curve X_0 forms a critical k-set whose count is unity and is not a critical curve of index $j \neq k$. For each set $F \neq S$ in $\mathfrak{F}$ containing X_0 there is a neighborhood N of X_0 relative to F containing no other homotopic critical curve.*

The proof of this result will be omitted. It can be established by argument like that used by Morse (XVII, p. 448) in the proof of a similar theorem.

THE UNIVERSITY OF CHICAGO,
CHICAGO, ILLINOIS.

REFERENCES.

I. H. Poincaré, "Sur les courbes définies par les équations différentielles," *Journal de Mathématiques*, 4th series, vol. 1 (1885), pp. 167-244.

II. G. D. Birkoff, "Dynamical systems with two degrees of freedom," *Transactions of the American Mathematical Society*, vol. 18 (1917), pp. 199-300.

III. M. Morse, "Relations between the critical points of a real function of n independent variables," *Transactions of the American Mathematical Society*, vol. 27 (1925), pp. 345-396.

IV. S. Lefschetz, "Continuous transformations of manifolds," *Proceedings of the National Academy of Sciences*, vol. 11 (1925), pp. 290-292.

V. G. D. Birkhoff, *Dynamical Systems*, American Mathematical Society Colloquium Publications, vol. 9 (1927).

VI. L. Vietoris, "Ueber den höheren Zusammenhang kompakter Raume und eine Klasse von zusammenhangstreuen Abbildungen," *Mathematische Annalen*, vol. 97 (1927), pp. 454-472.

VII. E. J. McShane, "Semicontinuity in the calculus of variations and absolute minima for isoperimetric problems," *Contributions to the Calculus of Variations*, 1930, University of Chicago Press.

VIII. L. Lusternik and L. Schnirelmann, "Méthodes topologiques dans les problèmes variationnels," *Actualités Scientifiques et industrielles*, vol. 188 (1934),

IX. M. Morse and Van Schaack, "The critical point theory under general boundary conditions," *Annals of Mathematics*, vol. 35 (1934), pp. 545-571.

X. M. Morse, *The Calculus of Variations in the Large*, American Mathematical Society Colloquium Publications, vol. 18 (1934).

XI. G. D. Birkhoff and M. R. Hestenes, "Generalized minimax principle in the calculus of variations," *Proceedings of the National Academy of Sciences*, vol. 21 (1935), pp. 96-99, *Duke Mathematical Journal*, vol. 1 (1935), pp. 413-432.

XII. M. Morse and van Schaack, "Abstract critical sets," *Proceedings of the National Academy of Sciences*, vol. 21 (1935), pp. 258-262.

XIII. S. Lefschetz, "On critical sets," *Proceedings of the National Academy of Sciences*, vol. 21 (1935), pp. 220-222, *Duke Mathematical Journal*, vol. 1 (1935), pp. 392-412.

XIV. M. R. Hestenes, "Minimax principle for functions," *Proceedings of the National Academy of Sciences*, vol. 22 (1936), pp. 115-119.

XV. M. Morse, "A special parametrization of curves," *Bulletin of the American Mathematical Society*, vol. 42 (1936), pp. 915-922.

XVI. M. Morse, *Analysis in the Large*. Lectures delivered at the Institute of Advanced Study 1936-7, Princeton University Press.

XVII. M. Morse, "Functional topology and abstract variational theory," *Annals of Mathematics*, vol. 38 (1937), pp. 386-449.

XVIII. H. Seifert and W. Threlfall, *Variationsrechnung im Grossen*, Hamburger Mathematische Einzelschriften, vol. 24 (1938), pp. 1-115.

XIX. M. Morse, *Functional topology and abstract variational theory*, Mémorial des Sciences Mathématiques, vol. 92 (1939), pp. 1-78.

XX. M. Morse, "Rank and span in functional topology," *Annals of Mathematics*, vol. 41 (1940), pp. 419-454.

XXI. W. Mayer, "A new approach to the critical value theory," *Bulletin of the American Mathematical Society*, vol. 46 (1940), pp. 838-847.

XXII. G. Ewing and M. Morse, "The variational theory in the large including the non-regular case—First paper," *Annals of Mathematics*, vol. 44 (1943), pp. 339-353.

XXIII. G. Ewing and M. Morse, "The variational theory in the large including the non-regular case—Second paper," *Annals of Mathematics*, vol. 44 (1943), pp. 354-374.

THE FOUR-VERTEX THEOREM FOR SURFACES OF CONSTANT CURVATURE.*

By S. B. JACKSON.

1. Introduction. The Four-Vertex Theorem, first published by Mukhopadhyaya in 1909 [1], states that every plane oval of class C'' possesses at least four extrema of the curvature, where an oval may be defined as a simple closed curve with non-vanishing curvature. This theorem has been extended to other classes of plane curves by Graustein [2], Fog [6], and the writer [4], and to certain limited classes of space curves [3, 7, 8]. In 1940 the writer extended the theorem in a slightly different direction by considering the extrema of the geodesic curvature on spherical curves [5]. The fundamental result of this paper was that every simple closed spherical curve of class C''', not a circle, has at least four extrema of the geodesic curvature. Such extrema are called geodesic vertices. The aim of the present discussion is to extend the results of this earlier paper to curves on any surface of constant Gaussian curvature. Specifically it will be shown (Theorem 4.1) that every simple closed curve of class C'', not a geodesic circle, in a simply connected region of a surface of constant curvature has at least four geodesic vertices.

The technique of the paper is to establish the existence of a transformation, called a transformation of type I, which maps the given simply connected region of the surface into the plane in such a way that the geodesic vertices of the given curve are carried into the vertices of the corresponding plane curve. The desired result then follows readily from the properties of this transformation and the known facts about the vertices of plane curves [4]. It is shown (**5**) by examples that the requirement that the simple closed curve lie in a simply connected region of the surface is essential.

No four-vertex theorem can be expected for surfaces of variable curvature since it is shown (**6**) that all sufficiently small distance circles [1] about a non-stationary point of the Gaussian curvature have only two geodesic vertices.

* Received February 2, 1945; Presented to the American Mathematical Society, April 27, 1945.

[1] A distance circle is what Blaschke [11, § 72] calls an Entfernungskreis, and is the locus of points at a fixed geodesic distance from a given point. On a surface of variable curvature distance circles are not geodesic circles as defined in this paper (**2**).

One type of stationary point is considered (7) and is characterized in terms of the geodesic polar coordinate system with center at this point.

2. Geodesic vertices and monotone arcs. A geodesic vertex of a curve on a surface, Σ, is a point or an arc of constant geodesic curvature for which the geodesic curvature is a relative extremum with respect to the neighboring arcs on either side. The term "relative extremum" is to be understood in the following sense. If $1/a$ is the geodesic curvature at the vertex and $1/\rho$ the geodesic curvature at an arbitrary point, then either $1/\rho - 1/a \geqq 0$ or $1/\rho - 1/a \leqq 0$ in a neighborhood of the vertex. The equality sign may hold at any number of points on the adjacent arcs but it must not hold at all points of either of them.

An arc for which the geodesic curvature is monotone non-decreasing or monotone non-increasing has no geodesic vertices, and conversely an arc with no geodesic vertices has this property. For brevity such an arc will be called monotone. All arcs and curves in the following discussion will be understood to have continuous geodesic curvature unless the contrary is stated.

The curves which have constant geodesic curvature are called geodesic circles.[2] We shall seek to characterize vertices and monotone arcs according to their behavior with respect to their osculating geodesic circles. It will be convenient to refer to points as lying to the left or right of a given directed curve. In such a statement it is understood that the surface is viewed from the tip of the directed normal to Σ and that the surface trihedral forms a right-handed system. Thus, in a coordinate system on Σ, the v-curves are always directed toward the left of the u-curves.

THEOREM 2.1. *Let C_1 and C_2 be curves of a surface Σ which are tangent in the same direction at p_0, and consider the correspondence established between the curves by the family of geodesics orthogonal to C_1. If, for all corresponding points sufficiently near p_0, $1/\rho_2 \geqq 1/\rho_1$, where $1/\rho_i$ is the geodesic curvature of C_i, then, in some neighborhood of p_0, C_2 lies on or to the left of C_1. The two curves either meet only at p_0 or coincide along a single arc, which necessarily contains p_0.*

Consider an orthogonal coordinate system on Σ with C_1 as the curve $u = 0$, and the geodesics orthogonal to C_1 as the u-curves. If u denotes arc length along these geodesics and v denotes arc length along C_1 measured in the positive direction from p_0, the first fundamental form is

$$ds^2 = du^2 + G\,dv^2,$$

[2] These are what Blaschke [11, § 72] calls Krümmungskreise as distinguished from Entfernungskreise. Geodesics are thus special geodesic circles.

where $G(0, v) = 1$. Since C_2 is tangent to C_1 at p_0, its equation can be written in the form $u = u(v)$, or equivalently, $\phi(u, v) \equiv u - u(v) = 0$, where $u(0) = u'(0) = 0$. The geodesic curvature of C_2 is readily found, by Bonnet's formula,[3] to be

$$(2.1) \quad 1/\rho_2 = [-2Gu'' + 2G_u u'^2 + G_v u' + GG_u]/[2\sqrt{G}(u'^2 + G)^{3/2}].$$

The geodesic curvature of C_1, whose equation is $u = 0$, is

$$(2.2) \qquad 1/\rho_1 = \bar{G}_u/2\bar{G},$$

where the bar denotes that the functions are evaluated for $u = 0$. The hypothesis of the theorem is that for sufficiently small v, $1/\rho_2 \geqq 1/\rho_1$. The result of substituting (2.1) and (2.2) into this relation may be written in the form

$$u'' \leqq G_u u'^2/G + G_v u'/2G + G_u/2 - (u'^2 + G)^{3/2}\bar{G}_u/2\bar{G}\sqrt{G}.$$

If the right hand side of the above inequality, which is formally a function of u since it contains $G(u, v)$ and its derivatives, is expanded in a finite Taylor series of two terms, the relation assumes the following form

$$u'' \leqq \bar{G}_v u'/2\bar{G} + [2u'^2/\bar{G} + 1 - (1 + u'^2/\bar{G})^{3/2}]\bar{G}_u/2 + Au$$

where the function A is bounded. The expression $(1 + u'^2/\bar{G})^{3/2}$ may be expanded in a convergent binomial series since, for small values of v, $|u'^2/\bar{G}| < 1$. If this is done the first two terms above have a common factor u'. By inserting absolute values on the right hand side, the inequality above thus implies that

$$(2.3) \qquad u'' \leqq K|u'| + L|u|,$$

where K and L are sufficiently large positive constants.

Recently R. P. Boas, Jr. has established the following result [9]. Let $f(x)$ be a real function possessing a second derivative, with an absolutely continuous first derivative, and satisfying the following two conditions.

i. $f(0) = f'(0) = 0$

ii. $f''(x) \leqq \phi(x)\,|f'(x)| + x^{-1}\psi(x)\,|f(x)|$ in the interval $[0, a]$

where $\phi(x)$ and $\psi(x)$ are positive and integrable. Then either $f(x) \equiv 0$ in some interval $[0, b]$ or $f(x) < 0$ in the interval $(0, \mathrm{Min}(a, c)]$ where $\int_0^c (\phi + \psi)\,dx = 1$.

[3] An inspection of the derivation of this formula [10, pp. 157-158] shows that the ambiguous sign is to be taken negative in this case.

This result applies immediately to the function $u(v)$ if we take $\phi = K$ and $\psi = Lu$. Thus either $u \equiv 0$ in some interval $0 \leqq v \leqq b$ or $u < 0$ in $0 < v \leqq b$, whence C_2 either coincides with C_1 in $[0, b]$ or lies to the left of it in $(0, b]$, with p_0 as the only common point for non-negative values of v.

For negative values of v, reverse the direction on the two curves, similarly reversing the coordinate system. This will change the sign of both curvatures, whence in this direction $1/\rho_1 \geqq 1/\rho_2$ or, equivalently, $-1/\rho_2 \geqq -1/\rho_1$. Substitution as above yields the inequality $-u'' \leqq K \mid u' \mid + L \mid u \mid$, from which we conclude as before that $-u \leqq 0$ or $u \geqq 0$. Thus when it is directed in this way C_2 lies on or to the right of C_1. As above the arcs from p_0 in this direction either meet only at p_0 or coincide along a single arc. This completes the proof of the theorem, since, as originally directed, C_2 lies on or to the left of C_1 in some neighborhood of p_0, and the two curves meet only at p_0 or coincide along a single arc which contains p_0 either as an endpoint or an interior point.

COROLLARY 2. 1. 1. *In the neighborhood of a geodesic vertex, a curve lies entirely to the left or entirely to the right of the osculating geodesic circle according as the geodesic curvature at the vertex is a minimum or a maximum.*

COROLLARY 2. 1. 2. *At an interior point of a monotone arc (or at an arc of constant geodesic curvature interior to a monotone arc) the arc crosses its osculating geodesic circle from right to left or from left to right according as the geodesic curvature is monotone non-decreasing or monotone non-increasing.*

Both corollaries are immediate consequences of the theorem. For Corollary 2. 1. 1 it is only necessary to take the curve and its osculating geodesic circle as the curves C_i. For Corollary 2. 1. 2, let one of the curves, C_i, be the path traced by a point which follows the arc as it approaches the vertex and follows the geodesic circle as it leaves it, while the other curve is the path obtained if the point follows the geodesic circle as it approaches the vertex and the arc as it leaves it.

LEMMA 2. 1. *A necessary and sufficient condition that an arc be monotone is that it cross every osculating geodesic circle at the point (or arc) of contact, the geodesic curvature being monotone non-decreasing or non-increasing according as the crossing is from right to left or from left to right.*

The necessity is precisely Corollary 2. 1. 2. The sufficiency follows from the fact that if the arc were not monotone there would be a vertex, which is impossible by Corollary 2. 1. 1. By exclusion we obtain at once the following characterization of geodesic vertices.

LEMMA 2. 2. *A point (or an arc of constant geodesic curvature) of a curve in the neighborhood of which the curve lies entirely on one side of the osculating geodesic circle is either a vertex or a limit point of vertices. If it is a vertex, the geodesic curvature is a maximum or a minimum according as the curve lies to the right or the left of the geodesic circle.*

It may be noted that if a point is a limit point of vertices there are infinitely many vertices of both kinds in any neighborhood of it.

3. Transformations of type *I*. A transformation of class C'' of a surface Σ to a surface Σ' is said to be of type I if it (a) is locally one-to-one, (b) preserves sense, and (c) carries geodesic circles into geodesic circles.

THEOREM 3. 1. *A transformation of type I carries arcs of monotone non-increasing (non-decreasing) geodesic curvature into arcs of monotone non-increasing (non-decreasing) geodesic curvature. It carries geodesic vertices into geodesic vertices of the same kind or into limit points of such vertices.*

This theorem is an immediate consequence of the definition of transformations of type I and the characterizations of monotone arcs and geodesic vertices contained in Lemma 2. 1, Lemma 2. 2, and Corollary 2. 1. 1.

In this discussion we shall be concerned entirely with the case where Σ' is the Gaussian plane, and this will be understood without further statement in the following development.

LEMMA 3. 1. *Every sufficiently small neighborhood of a point on a surface of constant Gaussian curvature, K, admits a one-to-one transformation of type I.*

If $K > 0$, a sufficiently small neighborhood of any point is applicable to a sphere of the same curvature, which in turn can be mapped into the plane by inversion. This transformation is of type I if the plane is suitably oriented. If $K = 0$, a neighborhood of any point is applicable to the plane, and this transformation is surely of type I. If, finally, $K < 0$, the familiar mapping of the neighborhood of a point into the Poincaré half plane is of type I [11, pp. 156-161], and the lemma is proved.

A neighborhood, on a surface of constant curvature, K, which admits a one-to-one transformation of type I into the plane will be termed a "special neighborhood." Throughout the remainder of the paper Σ will denote a surface of constant curvature unless the contrary is stated.

LEMMA 3. 2. *The one-to-one transformations of type I of a region R*

in a special neighborhood of Σ are obtained as the product of any particular one and the general direct circular transformation of the plane, $w = (az + b)/(cz + d)$, $(ad - bc \neq 0)$.

Such transformations are all clearly of type *I*. Suppose now that T_2 is any one-to-one transformation of type *I* of *R*, and T_1 is a second such transformation. Then $T_2T_1^{-1}$ is a transformation of a region of the plane into a region of the plane which carries circles and lines into circles and lines and which preserves sense. It is well known that the only such transformations are the direct circular transformations. Thus $T_2T_1^{-1} = F$ and $T_2 = FT_1$, where F is a direct circular transformation, and the proof is complete.

LEMMA 3.3. *A one-to-one transformation of type I of a region R in a special neighborhood of Σ is uniquely determined by three pairs of corresponding points.*

This is an immediate consequence of Lemma 3.2 and the fact that a direct circular transformation is a complex linear transformation and is thus uniquely determined by three pairs of corresponding points.

LEMMA 3.4. *If two transformations of type I (not necessarily one-to-one) on a region R of Σ coincide on a neighborhood of R, they coincide for all points of R.*

For otherwise there would be a point p of *R*, every neighborhood of which would have infinitely many points where the transformations coincide, and also points where they do not. By taking a special neighborhood of this point where both transformations are one-to-one, a contradiction to Lemma 3.3 is obtained.

LEMMA 3.5. *The transformations of type I (if any) of a region R on Σ are the products of any one such transformation and the general direct circular transformation.*

For in a special neighborhood where two transformations are one-to-one, one transformation is the product of the other by a direct circular transformation by Lemma 3.2, whence by Lemma 3.4 the relation is true for all points of *R*.

LEMMA 3.6. *All the transformations of type I of any region R of a special neighborhood N are one-to-one.*

For a transformation of type I on R is one-to-one in a sufficiently small neighborhood of any point p. By Lemma 3.3 there is a one-to-one transformation of type I for all N which coincides with the given transformation near p. By Lemma 3.4 the transformations then coincide on R and thus the given transformation is one-to-one on R.

LEMMA 3.7. *If R_1 and R_2 are two regions in a special neighborhood N of Σ, and if R_1 is contained in R_2, a one-to-one transformation of type I on R_1 uniquely determines a transformation of type I on R_2 which must also be one-to-one.*

Three points from R_1 together with their plane images uniquely determine a one-to-one transformation of type I for all of N, which coincides with the given transformation on R_1 by Lemma 3.3. Thus there is defined on R_2 a one-to-one transformation of type I which coincides with the given one on R_1. The uniqueness of the transformation follows from Lemma 3.4.

LEMMA 3.8. *Let S' be any closed set interior to a simply connected region S of a surface. Then, for any positive number ϵ, S' can be covered by a finite sequence, $C_1, \cdots, C_m$ of simply connected sub-regions satisfying the following conditions.*

(i) *The diameter of every C_i is less than ϵ.*

(ii) $\sum_{i=1}^{i=k} C_i$ *is simply connected,* $1 \leqq k \leqq m$.

(iii) *The intersection of C_{k+1} with $\sum_{i=1}^{i=k} C_i$ is simply connected for $1 \leqq k \leqq m-1$, and is contained in the union of a suitably chosen intersecting pair of the preceding C_i.*

Since S is simply connected, it is, by definition, the topological image of an open square, R, i. e. the correspondence is one-to-one and continuous in both directions. If R' is the set corresponding to S', then R' is closed and thus is contained interior to some closed square R'' in R. Since R'' is closed, the transformation of R'' into S is uniformly continuous, so that for any positive ϵ there exists a $\delta(\epsilon)$ such that if two points of R'' have a distance less than δ their images in S have distance less than ϵ.

Let R'' be divided into a network of squares by equally spaced lines parallel to the sides so that the diagonals of the small squares are less than $\delta/2$, and let the vertices of the net which are interior to R'' be ordered starting at the upper left hand corner and taking in order the points of the first row, then

the second row and so on. Let these points be $p_1, \cdots, p_m$. Corresponding to each p_i consider the open square C'_i whose center is p_i and which consists of the four small squares of the net meeting at p_i. These squares all have diagonals less than δ. This sequence of squares C'_i clearly satisfies conditions (ii) and (iii) above, for the regions obtained by adding the squares in order are all surely simply connected, and the intersection of any square with the previous ones consists of its intersection with the square above and the square on its left or either alone. Let C_i be the image of C'_i under the transformation. Conditions (ii) and (iii) are satisfied by the C_i since these properties are topological, while condition (i) is satisfied by construction. This completes the proof of the lemma.

THEOREM 3.2. *There exists a transformation of type I, not necessarily one-to-one, taking any simply connected region S of a surface of constant curvature Σ into a region of the plane.*

Consider a sequence of closed regions S_n $(n = 1, 2, \cdots)$, contained in S such that $S_{i+1} \supset S_i$ and $\sum_{i=1}^{\infty} S_i = S$. The images of a sequence of closed squares like R'' of Lemma 3.8 would serve the purpose. Let S_k be any one of these regions. Every point of S_k has a special neighborhood, and by the Heine-Borel Theorem a finite number of these neighborhoods N_i $(i = 1, \cdots, r)$, cover S_k. For any point p of S_k let $\rho(p)$ be half the radius of the largest distance circle [1] about p contained entirely in the closure of some N_i. The function $\rho(p)$ is continuous on the closed set S_k, and thus takes on its smallest value which is some positive number b. Thus the circle of radius b about any point of S_k is a special neighborhood.

Let $C_1, \cdots, C_m$ be the finite sequence of regions whose existence was established in Lemma 3.8, where ϵ is taken as b and the closed set S' as S_k. Any two intersecting C_i are thus contained in a special neighborhood since they are contained in a circle of radius b about any point of their intersection. Since C_1 is a region of a special neighborhood, it has a one-to-one mapping of type I into the plane. We shall show that this mapping can be extended successively to all the other C_i in such a way as to be one-to-one on each C_i. The proof is by induction. Suppose the mapping has been extended to $\sum_{i=1}^{i=r} C_i$, $1 \leqq r < m$ and consider C_{r+1}. The intersection of C_{r+1} with $\sum_{i=1}^{i=r} C_i$, by Lemma 3.8, is connected and is contained in $C_j \cup C_k$, where C_j and C_k are a suitably chosen intersecting pair of the preceding C_i. It may happen that $C_j = C_k$. Since $C_j \cup C_k$ is contained in a special neighborhood, the given mapping must

be one-to-one on $C_j \cup C_k$ by Lemma 3. 6. The transformation of type I is therefore one-to-one on the region common to C_{r+1} and $\sum_{i=1}^{i=r} C_i$. Inasmuch as C_{r+1} is contained in a special neighborhood, Lemma 3. 7 shows that the mapping of type I can be extended to all of C_{r+1} and is one-to-one on C_{r+1}. Thus the mapping is extended to $\sum_{i=1}^{i=r+1} C_i$ and the induction is complete. There exists therefore a transformation of type I on any S_k. It will be noted that in this discussion the terms transformation and mapping have been used interchangeably.

Let T_1 be any transformation of type I on S_1. We shall show that this transformation can be extended to all S_n. It has just been shown that there is a transformation T_2 of type I on S_2. Since T_1 and T_2 are both transformations of type I on S_1, it follows from Lemma 3. 5 that there is a direct circular transformation F such that $T_1 = FT_2$ on S_1. The transformation FT_2 is thus a transformation of type I on S_2 which coincides with the given transformation on S_1. The given transformation has thus been extended to S_2. Proceeding by induction, the transformation can now be extended to all S_k. But since every point of S is in some S_k, the transformation can be extended to all points of S and the proof is complete.

4. Transformation of type *I* of a closed curve on Σ. Let C be a closed differentiable curve in a simply connected region S of Σ, and let S be mapped into the plane by a transformation of type I as is possible by Theorem 3. 2. Let S' denote the closed set of points consisting of C and the simply connected regions of S bounded by C, and let $\bar{S}$ be the set of points in the plane into which S' is carried by the transformation. In a sufficiently small neighborhood of any point p of S' the transformation of S' to the corresponding points of $\bar{S}$ is one-to-one. The transformation need not be one-to-one in the large since a point in $\bar{S}$ may have several inverses in S'.

LEMMA 4. 1. *A point of $\bar{S}$ has at most a finite number of inverses in S'.*

For if any point $\bar{p}$ in $\bar{S}$ had an infinite number of inverses, then by the Weierstrass-Bolzano Theorem the inverses would have a limit point p which, by continuity, would also be an inverse of $\bar{p}$. But then every neighborhood of p would contain points which map into $\bar{p}$ and the transformation would not be one-to-one near p. Since this is impossible the lemma is proved.

LEMMA 4. 2. *In a sufficiently small neighborhood of any point $\bar{p}$ of $\bar{S}$ there are no points with more inverses in S' than $\bar{p}$.*

Let the number of inverses of p be k, which is finite by Lemma 4.1. If the lemma is false there is a sequence of points $\bar{p}_i$ converging to $\bar{p}$, each of which has at least $k+1$ inverses. Consider the set of limit points of these inverses. By continuity these are all inverses of $\bar{p}$. About each of the k inverses of $\bar{p}$ let there be taken a neighborhood so small that no two points have the same image in $\bar{S}$. All but a finite number of the inverses of all $\bar{p}_i$ lie in these k neighborhoods. Thus some $\bar{p}_j$ has at least $k+1$ inverses in these k neighborhoods, whence it has two inverses in some one neighborhood. Since this is false by construction we have a contradiction and the lemma is proved.

LEMMA 4.3. *If a point $\bar{p}$ of $\bar{S}$ is not the image of a point of C, then there is a neighborhood of $\bar{p}$ every point of which has the same number, k, of inverses. These inverses make up neighborhoods of the k inverses of $\bar{p}$ and the mapping of each of these non-intersecting neighborhoods into the one about $\bar{p}$ is one-to-one and thus topological.*

Let the k inverses of $\bar{p}$ be p_i, $1 \leqq i \leqq k$, none of which lie on C. About each p_i take a neighborhood N'_i which is a special neighborhood, which has no points in common with C so that N'_i is interior to S', and so that no two N'_i have a point in common. The mapping of each N'_i into a neighborhood $\bar{N}'_i$ of $\bar{p}$ is one-to-one. Take a neighborhood $\bar{N}$ of $\bar{p}$ interior to all the $\bar{N}'_i$ and so that all points of $\bar{N}$ have not more than k inverses, as is possible by Lemma 4.2. Every point of $\bar{N}$ has then exactly k inverses since there is one in each of the non-overlapping N'_i. If N_i denotes the set of points of N'_i which go into $\bar{N}$, these neighborhoods satisfy the conditions of the lemma and the proof is complete.

Let the plane curve which is the image of C be denoted by $\bar{C}$. The curve $\bar{C}$ divides the plane into a certain number of finite regions and one infinite one (it is always possible to choose the mapping so that $\bar{C}$ does not contain the point at infinity itself). Considered in the Gaussian plane the infinite region is simply connected, as are all the finite ones, and in the following discussions we shall refer to them all as the simply connected regions of the plane bounded by $\bar{C}$.

LEMMA 4.4. *The number of inverses, k, is the same for all points of one of the simply connected regions bounded by the image curve $\bar{C}$.*

By Lemma 4.3 the number of inverses in such a region is a continuous, integral valued function and is therefore constant over the region. This establishes the result.

Let $\bar{R}$ be any one of the simply connected regions into which $\bar{C}$ divides

the plane. The inverse set of $\bar{R}$ in S' is clearly open both in S and S'. Let R be any one of the components of this inverse set. Since on the open set R the transformation is open, its image is an open connected subset of $\bar{R}$, say $\bar{Q}$. Suppose that $\bar{Q}$ has a boundary point $\bar{p}$ interior to $\bar{R}$. Take a neighborhood $\bar{N}$ of $\bar{p}$ interior to $\bar{R}$ which has the property established in Lemma 4. 3. Then since $\bar{N}$ intersects $\bar{Q}$, some one of the inverse neighborhoods N intersects R. Thus $N + R$ is an open connected subset of S' mapping into $\bar{R}$. But this is impossible for R was a component of the inverse set and hence already maximal. Thus no boundary point of $\bar{Q}$ is interior to $\bar{R}$. But the only connected subset of $\bar{R}$ bounded entirely by boundary points of $\bar{R}$ is $\bar{R}$ itself, whence $\bar{Q} = \bar{R}$, and the component R maps into the whole of $\bar{R}$.

By Lemma 4. 3 every point $\bar{p}$ of $\bar{R}$ has a neighborhood which is the topological image of each of the k non-overlapping inverse neighborhoods. If $\bar{T}$ is any closed subset of $\bar{R}$, it follows readily from the Heine-Borel Theorem that there exists an ϵ such that the ϵ-neighborhood of every point $\bar{p}$ of $\bar{T}$ has the property stated above. Now let $\bar{q}$ be any point of $\bar{R}$ and q any one of its inverses in R. If a moving point $\bar{p}$ traces out a path in $\bar{R}$ starting from $\bar{q}$, there is defined by continuity a unique path in R starting from q, for by the above discussion whenever the mapping has reached a point $\bar{t}$ of the path it can be extended throughout the ϵ-neighborhood of $\bar{t}$ and thus does not terminate. In order to apply the argument it is only necessary to take $\bar{T}$ as any closed set in $\bar{R}$ containing the path.

Finally we shall show that the transformation of R into $\bar{R}$ is one-to-one. Let $\bar{q}$ be any point of $\bar{R}$ and q_1 and q_2 any two of its inverses in R. Let q_1 and q_2 be joined by a curve in R. The image curve in $\bar{R}$ is therefore closed. Let this closed curve be drawn together to the point $\bar{q}$, which is possible since $\bar{R}$ is simply connected. Each point of this curve will move continuously over a path in $\bar{R}$, and thus this operation will, by the results of the last paragraph, generate a continuous deformation of the curve q_1q_2 into a point. But this is impossible unless $q_1 = q_2$, whence every point of $\bar{R}$ has exactly one inverse in R, and the transformation is one-to-one on R.

The same argument applies to each of the other components of the inverse of $\bar{R}$ in S'. There are clearly k of these components, since each one contains one inverse of each point. Since the transformation of any of these components R on $\bar{R}$ has been proved one-to-one and continuous, it is topological. In view of this topological character of the transformation, we can conclude that the components are all simply connected, like $\bar{R}$. The results of the last few paragraphs may be summarized in the following lemma.

LEMMA 4. 5. *The inverses of any one of the simply connected regions* $\bar{R}$

of the plane bounded by $\bar{C}$ are a finite set of disjoint simply connected regions, each of which is mapped topologically on $\bar{R}$ by the given transformation of type I.

LEMMA 4. 6. *The boundary of any one of the k inverses, R_j, of a simply connected region, $\bar{R}$, of the plane bounded by $\bar{C}$ maps into the boundary of $\bar{R}$.*

Any boundary point, b, of R_j can be approached by a sequence of points of R_j, whose images in $\bar{R}$ converge to the image of b. If this image were interior to $\bar{R}$ the point b would necessarily be an interior point of some other R_i, whence R_i and R_j would overlap which is impossible by Lemma 4. 5. Conversely, every boundary point of $\bar{R}$ can be approached by a sequence of points in $\bar{R}$ whose inverses in R_j have one or more limit points in S'. Any such limit point, by the continuity of the transformation, maps into the given boundary point of $\bar{R}$. If such a limit point were interior to R_j it would map into the interior of $\bar{R}$. Since this is impossible, all these limit points are boundary points of R_j. It should be clearly noted that the mapping on the boundary need not be one-to-one.

An arc AB of a closed plane curve $\bar{C}$ is called a simple loop if (a) it is simple except that $A = B$, and (b) one of the regions into which it divides the plane contains no further points of the curve. This region is then one of the simply connected regions of the plane bounded by $\bar{C}$, and is referred to as the region corresponding to the loop.

Up to this point C has been any closed curve of class C'' in S. Now let C be in particular a simple closed curve, positively directed, i. e. with its interior to its left.

LEMMA 4. 7. *If C is a positively directed simple closed curve of class C'' in the simply connected region S of Σ, then under a transformation of type I the corresponding plane curve $\bar{C}$ can contain no simple loop whose corresponding region lies to its left.*

Suppose that the lemma is false. Then there is a simple loop whose region $\bar{R}$ lies to its left. This region may be assumed finite since in any case it can be made so by a direct circular transformation. Since $\bar{C}$ is the image of C, there is some arc ab of C which maps into the loop, where both a and b map into the double point of the loop. The points close to arc ab and to its left map into the points close to the loop and to its left, i. e. into $\bar{R}$, since the transformation preserves sense. Thus ab is a part of the boundary of one of the inverse regions R_i of $\bar{R}$. Since ab is not closed it cannot be the entire boundary of R_i. Let p be any other point of the boundary of R_i which does

not map into the double point of the loop. Such points p exist, since ab and all the inverses of the double point together still do not bound a region in S. By Lemma 4. 6, p maps into a point $\bar{p}$ of the loop. Take a neighborhood of p in S so small that the mapping in this neighborhood is topological. This neighborhood may include points of S not in S' and the transformed neighborhood at $\bar{p}$ may include points of the plane not in $\bar{S}$. Take a smaller neighborhood $\bar{N}$ if necessary about $\bar{p}$ which is divided in exactly two parts by the loop. This is possible since the loop is differentiable. Since the mapping of $\bar{N}$ on its inverse neighborhood N at p is topological, N is divided into exactly two half neighborhoods by the inverse of the arc. Some point in one of these half neighborhoods is in R_i, and by hypothesis this maps into $\bar{R}$, i. e. to the left of the directed loop at $\bar{p}$. All points of this half neighborhood are thus in R_i since any two can be joined by a path whose image lies entirely in $\bar{R}$. But there is also a point p' of ab which maps into $\bar{p}$, and the set of points to the left of ab near p' also maps into the points of $\bar{R}$ near $\bar{p}$ and to the left of the loop. Thus a point of $\bar{R}$ near $\bar{p}$ has one inverse in R_i near p and a second near p'. This is impossible, since by Lemma 4. 5 the transformation of R_i on $\bar{R}$ is one-to-one. Thus we have a contradiction, and the lemma is proved.

THEOREM 4. 1. *Every simple closed curve of class C'', not a geodesic circle, in a simply connected region of a surface of constant curvature Σ has at least four geodesic vertices.*

Let the simple closed curve C be mapped on the plane by a transformation of type I as discussed above. By Theorem 3. 1 the geodesic vertices of C go into the vertices of the plane curve $\bar{C}$, and by Lemma 4. 7 $\bar{C}$ has no simple loop having its corresponding region to its left. But it is known [4, Theorem 5. 1] that a plane curve having only two vertices has two simple loops, one of which has the corresponding region to its left. Thus $\bar{C}$ cannot have only two vertices, whence $\bar{C}$ and therefore C has at least four geodesic vertices and the proof is complete.

It may be observed that this discussion establishes the Four-Vertex Theorem for a very large number of curves which may, at first glance, appear to be outside the class under consideration. Many plane curves for example are not simple but become simple when viewed as lying on a many sheeted covering surface of the plane, and are thus covered by the above theorem by taking the covering surface for the surface Σ. Similar remarks apply to curves on other surfaces, such as cylinders and cones, which are not simple on these surfaces but are simple on some covering surface.

5. Counterexamples. This section is devoted to two examples showing that the restriction in Theorem 4. 1 that C lie in a simply connected region of Σ cannot be dispensed with.

Example 1.[4] Consider, as the surface Σ, the right circular cylinder with generators parallel to the z-axis and with the circle $x = a\cos\theta$, $y = a\sin\theta$ in the xy-plane as directing curve, and let the curve C be the intersection of Σ with the plane $z = by$. If s denotes the arc length on the directing circle measured from the x-axis, then $s = a\theta$. For the points of C we have therefore $z = by = ba\sin\theta = ba\sin(s/a)$, where $0 \leqq s < 2\pi a$. If the cylinder is rolled out on a plane, C goes into the curve, $y' = ba\sin(x'/a)$, $0 \leqq x' < 2\pi a$. But in one period a sine curve has exactly two vertices, whence the same is true of C. C is thus a simple closed curve with only two geodesic vertices on a surface of constant curvature. It will be noted that C cannot be deformed into a point and thus is not in a simply connected region of Σ. This shows that in general it is not possible to eliminate the requirement that C lie in a simply connected region of Σ, if Theorem 4. 1 is to hold.

Example 2. This example will show furthermore that if Theorem 4. 1 is to be valid it is not even possible to allow a singularity of the surface in the simply connected region S. Let the surface Σ be a right circular cone with its vertex at $(0, 0, h)$ and the circle $x = a\cos\theta$, $y = a\sin\theta$ in the xy-plane for directing curve, and let C be the intersection of this cone with the plane $z = by$, where $b < h/a$ so that the plane actually cuts all the generators of the lower nappe of the cone. C is then a simple closed curve on Σ with a singular point, namely the vertex, in its interior. If (r, θ, z) are cylindrical coordinates of a general point of C, the following relations are immediate.

$$(5.1) \qquad \begin{aligned} &x = r\cos\theta, \qquad && y = r\sin\theta \\ &r/a + z/h = 1. \end{aligned}$$

Let s have the same meaning as in Example 1, l be the slant height of the cone from the vertex to the xy-plane, ϕ be defined as s/l, and ρ be the distance from the vertex to the general point of C. It is easy to verify that

$$(5.2) \qquad \begin{aligned} &l^2 = a^2 + h^2, \qquad && s = a\theta \\ &\phi = (a/l)\theta, \qquad && \rho = (l/a)r. \end{aligned}$$

Substitution from equations (5. 1) and (5. 2) into the relation $z = by$ yields the following relation between ρ and ϕ, for the equation of C:

---

[4] This example was first suggested to the writer by the late W. C. Graustein. It was also suggested by the referee.

$$(5.3) \qquad \rho = \frac{hl}{h + ab \sin (l\phi/a)}, \qquad h > ab, \qquad 0 \leqq \phi < 2\pi a/l.$$

If the cone is cut along the line $\theta = 0$ and flattened out on a plane, the quantities ρ and ϕ of equation (5.3) become precisely the polar coordinates of the points on the corresponding plane curve, and since geodesic curvature is preserved by the operation, it is only necessary to seek the vertices of the plane curve having (5.3) as its polar coordinate equation. Straightforward computation shows that the curvature is

$$(5.4) \quad K = \frac{[a - bh \sin (l\phi/a)][h + ab \sin (l\phi/a)]^3}{al[h^2 + a^2b^2 + b^2h^2 + 2abh \sin (l\phi/a) - b^2h^2 \sin^2 (l\phi/a)]^{3/2}}.$$

Consider now the numerical case $a = 1$, $b = 1/2$, $h = 2$. Equation (5.4) becomes

$$(5.5) \qquad K = \frac{(1 - \sin \sqrt{5}\phi)(4 + \sin \sqrt{5}\phi)^3}{\sqrt{5}(21 + 8 \sin \sqrt{5}\phi - 4 \sin^2 \sqrt{5}\phi)^{3/2}}.$$

If $dK/d\phi$ is equated to zero to find the extreme values of K, it appears that ϕ must be a solution of one or the other of the following equations.

$$(5.6) \qquad \begin{aligned} &4 \cos^3 \sqrt{5}\phi - 52 \cos^2 \sqrt{5}\phi - 8 \cos \sqrt{5}\phi - 69 = 0 \\ &\cos \sqrt{5}\phi = 0. \end{aligned}$$

The first of equations (5.6), considered as a cubic equation in $\cos \sqrt{5}\phi$, has only one real root and this is bigger than 1, so no value of ϕ will satisfy this equation. The second equation has only two solutions in the interval $0 \leqq \phi < 2\pi/\sqrt{5}$, whence C has only two geodesic vertices. Thus the conclusion of Theorem 4.1 is invalid for the curve C.

It is of some interest to note, in connection with Example 2, that not all the curves C obtained in the manner described have just two vertices. For example the curve C obtained by taking $a = 1$, $b = 1/2$, $h = 1$ turns out to have six geodesic vertices, as may be readily verified by carrying out the computation as indicated above.

6. Surfaces of variable curvature. In this section it will be shown that Theorem 4.1 cannot be extended to surfaces of variable Gaussian curvature. Specifically we shall show that on such surfaces sufficiently small distance circles [1] have only two geodesic vertices.

Let o be any point of a surface, Σ, of variable Gaussian curvature, and let (u, v) be geodesic polar coordinates on Σ with origin at o. The geodesic curvature of the v-curves, that is, of the distance circles about o, is given [10, p. 149] by

$$(6.1) \qquad 1/\rho = (1/\sqrt{G})(\partial\sqrt{G}/\partial u).$$

We seek the zeros of $\partial(1/\rho)/\partial v$ for $u = \text{const.} \neq 0$, as v varies from 0 to 2π, for the geodesic vertices of this distance circle are precisely the points where $\partial(1/\rho)/\partial v$ changes sign. After differentiation (6.1) becomes

$$(6.2) \qquad \frac{\partial}{\partial v}\left(\frac{1}{\rho}\right) = \frac{1}{\sqrt{G}}\frac{\partial^2\sqrt{G}}{\partial u \partial v} - \frac{1}{G}\frac{\partial\sqrt{G}}{\partial u}\frac{\partial\sqrt{G}}{\partial v}.$$

Since $G = 0$ only at $u = 0$, it suffices to find the zeros of

$$(6.3) \qquad G\frac{\partial}{\partial v}\left(\frac{1}{\rho}\right) = \sqrt{G}\frac{\partial^2\sqrt{G}}{\partial u \partial v} - \frac{\partial\sqrt{G}}{\partial u}\frac{\partial\sqrt{G}}{\partial v}.$$

Assuming sufficient differentiability for Σ, we may expand $\sqrt{G}$ in a power series in u. It is known [10] that $(\sqrt{G})_{(0,v)} = 0$, $(\partial\sqrt{G}/\partial u)_{(0,v)} = 1$. The other coefficients in the expansion may be computed from these by using the Gauss equation, $\partial^2\sqrt{G}/\partial u^2 = -K\sqrt{G}$, and its derivatives at $u = 0$. The subscripts above indicate evaluation for $u = 0$ but v arbitrary. By this method we obtain

$$(6.4) \qquad \sqrt{G} = u - \tfrac{1}{6}K_{(0,v)}u^3 - \tfrac{1}{12}(\partial K/\partial u)_{(0,v)}u^4 + \tfrac{1}{120}(K^2 - 3\partial^2 K/\partial u^2)_{(0,v)}u^5 + \cdots.$$

Differentiation of (6.4) and substitution in (6.3) yields the result

$$G\partial(1/\rho)/\partial v = -\tfrac{1}{4}(\partial^2 K/\partial u \partial v)_{(0,v)}u^4 - \tfrac{1}{10}(\partial^3 K/\partial u^2 \partial v)_{(0,v)}u^5 + \cdots.$$

Since, for our purpose, $u \neq 0$, it is enough to find the zeros of $\phi(u, v)$, where

$$(6.5) \qquad \phi(u, v) = \tfrac{1}{4}(\partial^2 K/\partial u \partial v)_{(0,v)} + \tfrac{1}{10}(\partial^3 K/\partial u^2 \partial v)_{(0,v)}u + \cdots.$$

for small fixed u and variable v.

Let C_1 and C_2 be two perpendicular curves on the surface Σ meeting at o, with arc lengths s_1 and s_2 respectively, and let θ be the angle from C_1 to a curve C through o. Precisely as in the case of the plane, the directional derivative of a function F on Σ in the direction of C at o is given by

$$\partial F/\partial s = (\partial F/\partial s_1)\cos\theta + (\partial F/\partial s_2)\sin\theta$$

where $\partial F/\partial s_i$ is the directional derivative of F in the direction of C_i. We shall apply this to the Gaussian curvature K on Σ. Let C_1 and C_2 be the curves $v = 0$ and $v = \pi/2$, respectively. Then $\partial F/\partial s_1 = (\partial K/\partial u)_{(0,0)}$, $\partial F/\partial s_2 = (\partial K/\partial u)_{(0,\pi/2)}$, $\partial F/\partial s = (\partial K/\partial u)_{(0,v)}$, $\theta = v$, and the above equation becomes

$$(6.6) \qquad (\partial K/\partial u)_{(0,v)} = (\partial K/\partial u)_{(0,0)}\cos v + (\partial K/\partial u)_{(0,\pi/2)}\sin v.$$

By differentiating (6.6) the following relation is obtained.

$$(6.7)\quad (\partial^2 K/\partial u\partial v)_{(0,v)} = -(\partial K/\partial u)_{(0,0)} \sin v + (\partial K/\partial u)_{(0,\pi/2)} \cos v.$$

If $(\partial^2 K/\partial u\partial v)_{(0,v)} \equiv 0$, then by (6.7) $(\partial K/\partial u)_{(0,0)} = (\partial K/\partial u)_{(0,\pi/2)} = 0$, whence, by (6.6), $(\partial K/\partial u)_{(0,v)} \equiv 0$, and o is a stationary point of K on Σ. Let o now be a non-stationary point of K. Such points exist, for otherwise Σ would have constant curvature. There are then exactly two values of v on the interval $0 \leqq v < 2\pi$ for which $(\partial^2 K/\partial u\partial v)_{(0,v)} = 0$, namely the solutions of the equation

$$(6.8)\qquad \tan v = \frac{(\partial K/\partial u)_{(0,\pi/2)}}{(\partial K/\partial u)_{(0,0)}}.$$

These solutions may be denoted by v_1 and v_2 where $v_2 = v_1 + \pi$.

Consider first a point $(0, \bar{v})$, where $\bar{v} \neq v_1, v_2$. Then $\phi(0, \bar{v}) = (1/4)(\partial^2 K/\partial u\partial v)_{(0,v)} \neq 0$. By continuity there exists a neighborhood of the coordinates $(0, \bar{v})$ for which $\phi(u, v) \neq 0$.[5] On the other hand consider the points $(0, v_i)$, $i = 1, 2$. By definition of v_i (6.8) holds, whence it follows that

$$(6.9)\qquad \sin v_i = \pm \frac{(\partial K/\partial u)_{(0,\pi/2)}}{m}, \qquad \cos v_i = \pm \frac{(\partial K/\partial u)_{(0,0)}}{m}$$

where $m^2 = (\partial K/\partial u)^2_{(0,0)} + (\partial K/\partial u)^2_{(0,\pi/2)} \neq 0$, since o is not a stationary point of K. Differentiation of (6.7) with respect to v and substitution from (6.9) yields the result

$$(6.10)\quad (\partial^3 K/\partial u\partial v^2)_{(0,v)} = -(\partial K/\partial u)_{(0,0)} \cos v - (\partial K/\partial u)_{(0,\pi/2)} \sin v$$
$$= \mp m(\cos v \cos v_i + \sin v \sin v_i) = \mp m \cos(v - v_i).$$

Hence $(\partial^3 K/\partial u\partial v^2)_{(0,v_i)} = \mp m \neq 0$, and

$$(6.11)\qquad \partial\phi(0, v_i)/\partial v \neq 0.$$

By (6.11) and the implicit function theorem, there exists a neighborhood of the coordinates $(0, v_i)$ [5] in which the equation $\phi(u, v) = 0$ can be solved for v uniquely and exhaustively. Hence to every point of the set $(0, v)$, $0 \leqq v \leqq 2\pi$, has been ordered a neighborhood of this set of coordinates. By the Heine-Borel Theorem there exists a finite set of these neighborhoods covering this set. A positive u_0 can then be chosen so small that (u_0, v) is always in one

[5] This is a neighborhood of the coordinate pair $(0, \bar{v})$, i.e., the set of pairs (u, v) such that $|u| < \delta$, $|v - \bar{v}| < \delta$. It is not a neighborhood of o on Σ.

of the neighborhoods selected. For such a u_0 there are only two zeros of $\phi(u_0, v)$. Hence $u = u_0$ has only two geodesic vertices. We have proved the following theorem.

THEOREM 6.1. *The distance circles of sufficiently small radius about a non-stationary point of the Gaussian curvature on a surface* Σ *have only two geodesic vertices.*

This result makes it clear that no four-vertex theorem can be expected on surfaces of variable Gaussian curvature.

Theorem 6.1 makes it possible to establish the following characterization of surfaces of constant curvature.

THEOREM 6.2. *A necessary and sufficient condition that a surface* Σ *be of constant curvature is that some neighborhood of every point admit a transformation of type I into the plane.*

The necessity is implied by Theorem 3.2. It remains to prove the sufficiency. Let p be a non-stationary point of the Gaussian curvature, and suppose a region containing p admits a transformation of type I. Since the transformation is locally one-to-one, a sufficiently small distance circle about p will be carried into a simple closed plane curve, which therefore has four vertices or is a circle. But by Theorem 3.1 this means that the distance circle has four vertices or none contrary to Theorem 6.1. This is a contradiction, whence every point is a point of stationary Gaussian curvature, and the surface has constant curvature.

7. Stationary points of the Gaussian curvature. Among the stationary points of the curvature mentioned in **6**, one type will be given further analysis. Let us consider a point o of a surface Σ for which all sufficiently small distance circles have constant geodesic curvature. Such points must have stationary Gaussian curvature, as was seen in **6**. A vertex of an ellipsoid of revolution, for example, is such a point.

Using the coordinate system of **6**, the condition that the distance circles about o have constant geodesic curvature is expressed by the equation $\partial(1/\rho)/\partial v \equiv 0$ for all v and all sufficiently small u, where $1/\rho$ is the geodesic curvature of the distance circles. By (6.1) this is equivalent to

$$\frac{\partial}{\partial v}\,\frac{1}{\sqrt{G}}\,\frac{\partial\sqrt{G}}{\partial u} = 0 \tag{7.1}$$

whence there follows

$$\frac{1}{\sqrt{G}}\frac{\partial\sqrt{G}}{\partial u} = 1/u + a(u) \tag{7.2}$$

where $a(u)$ has no singularity near $u = 0$. The form of the function on the right side follows from (6.4). By integrating (7.2) we obtain the following equation

$$\sqrt{G} = b(v)\exp\{\int[1/u + a(u)]du\} = ub(v)\exp[A(u)] \tag{7.3}$$

where $A(u) = \int a(u)du$. As in **6**, $(\partial\sqrt{G}/\partial u)_{(0,v)} = 1$, whence it appears that $b(v) = \exp[-A(0)]$. From this there follows

$$\sqrt{G} = u\exp[A(u) - A(0)]; \tag{7.4}$$

that is, G is a function of u alone. Moreover condition (7.4) is sufficient as well as necessary, for a $\sqrt{G}$ of form (7.4) satisfies the conditions for geodesic polar coordinates and certainly satisfies (7.1). We have thus established the following theorem.

THEOREM 7.1. *A necessary and sufficient condition that all distance circles of sufficiently small radius about a point o of a surface* Σ *have constant geodesic curvature is that the third coefficient, G, of the first fundamental form for the geodesic polar coordinate system about o be a function only of the geodesic distance from o.*

COROLLARY 7.1.1. *A necessary and sufficient condition that all distance circles of sufficiently small radius about a point o of a surface* Σ *have constant geodesic curvature is that these distance circles and the geodesics through o form an isometric system.*

The corollary follows at once from Theorem 7.1 and the conditions that the parametric curves form an isometric system [10, p. 133], namely $F = 0$, and $E/G = U(u)/V(v)$.

UNIVERSITY OF MARYLAND.

BIBLIOGRAPHY.

1. S. Mukhopadhyaya, "New methods in the geometry of a plane arc," *Bulletin of the Calcutta Mathematical Society*, vol. 1 (1909), pp. 31-37.

2. W. C. Graustein, "Extensions of the four-vertex theorem," *Transactions of the American Mathematical Society*, vol. 41 (1937), pp. 9-23.

3. W. C. Graustein and S. B. Jackson, "The four-vertex theorem for a certain type of space curves," *Bulletin of the American Mathematical Society*, vol. 43 (1937), pp. 737-741.

4. S. B. Jackson, "Vertices of plane curves," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 564-578.

5. ———, "The four-vertex theorem for spherical curves," *American Journal of Mathematics*, vol. 62 (1940), pp. 795-812.

6. D. Fog, "Über den Vierscheitelsatz und seine Verallgemeinerungen," *Sitzungsberichte der Berlin Akademie der Wissenschaft* (1933), pp. 251-254.

7. W. Süss, "Ein Vierscheitelsatz bei geschlossenen Raumkurven," *The Tôhoku Mathematical Journal*, vol. 29 (1928), pp. 359-362.

8. T. Takasu, "Vierscheitelsatz für Raumkurven," *The Tôhoku Mathematical Journal*, vol. 39 (1934), pp. 292-298. There are also several other papers by the same author on this general subject.

9. R. P. Boas, Jr., "A differential inequality," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 95-96.

10. W. C. Graustein, *Differential Geometry*, Macmillan (1935).

11. W. Blaschke, *Vorlesungen über Differentialgeometrie* I, Springer (1930).

CONFORMAL CLASSES OF SURFACES.*

By AARON FIALKOW

Introduction. In previous papers [1] [2, 3, 4], we have developed a method for the study of those properties of a subspace V_n $(1 \leqq n \leqq m-1)$ imbedded in any enveloping space V_m which remain unchanged under conformal transformations of V_m. The pictorial interpretation of the analytic apparatus described in these papers will appear elsewhere. The present paper applies some of these general rsults to a discussion of various classes of surfaces V_2 in Euclidean 3-space R_3 which have a simple conformal character. The results also apply to conformally equivalent surfaces in any conformally Euclidean 3-space. The reader is referred to [4] for the notation used as well as for a statement of the assumptions concerning the class of the functions which occur in this paper.

2. The analytic apparatus. The results stated below are a summary of some of the analytic aspects, used in this paper, of the conformal theory of a hypersurface V_n $(n > 1)$ in R_{n+1}. The details of this theory for the general case of a V_n in any V_m and the proofs appear in [4]. Let $a_{\alpha\beta}$, $\left\{ \begin{matrix} \alpha \\ \beta\gamma \end{matrix} \right\}$ be the metric tensor and Christoffel symbols of the second kind of a Euclidean space R_{n+1} in some coördinate system $\{y^\alpha\}$. Let

$$y^\alpha = y^\alpha(x^1, x^2, \cdots, x^n) \tag{2.1}$$

be the equations of a V_n imbedded in R_{n+1}. Let g_{ij}, $\left\{ \begin{matrix} i \\ jk \end{matrix} \right\}$ be the induced metric tensor and the Christoffel symbols of the second kind of V_n in the parameter system $\{x^i\}$. Covariant differentiation of (2.1) with respect to $\left\{ \begin{matrix} \alpha \\ \beta\gamma \end{matrix} \right\}$, $\left\{ \begin{matrix} i \\ jk \end{matrix} \right\}$, denoted by a comma, gives $y_{,ij}{}^\alpha = b_{ij}\zeta^\alpha$, where the b_{ij} are the coefficients of the second fundamental form of V_n and ζ^α is the unit normal to V_n. The mean curvature normal μ^α of V_n is defined by

$$\mu^\alpha = 1/n \cdot y_{,ij}{}^\alpha g^{ij}. \tag{2.2}$$

Its length is the mean curvature μ so that

* Received June 20, 1944.

[1] The numbers refer to the bibliography at the end of the paper.

$$\mu = (1/n)b_{ij}g^{ij} = (1/n)\sum_{i=1}^{n} {}_{(i)}b.$$

The ${}_{(i)}b$ are the n principal normal curvatures of V_n.

Let $\{\bar{y}^a\}$ be another coördinate system in R_{n+1} and suppose that the point transformation $\bar{y}^a = y^a$ is a conformal mapping of R_{n+1} on itself. Since the differential elements of distance at corresponding points of R_{n+1} are related by the equation

$$d\bar{s} = e^{\sigma} ds \tag{2.3}$$

it follows that $\bar{a}_{\alpha\beta} = e^{2\sigma}a_{\alpha\beta}$. The conformal mapping $\bar{V}_n \leftrightarrow V_n$ induced by the parametric equations $\bar{x}^i = x^i$ is such that $\bar{g}_{ij} = e^{2\sigma}g_{ij}$. The tensor $G_{ij}{}^a$ defined by

$$G_{ij}{}^a = y_{,ij}{}^a - g_{ij}\mu^a \tag{2.4}$$

is unchanged by conformal transformations of R_{n+1}. The *relative conformal curvature* e^{ϕ} is defined by

$$e^{2\phi} = (1/n)a_{\alpha\beta}g^{hi}g^{jk}G_{hj}{}^{\alpha}G_{ik}{}^{\beta} = (1/n)\sum_{i=1}^{n} ({}_{(i)}b - \mu)^2. \tag{2.5}$$

It is the standard deviation of the statistical population consisting of the n principal normal curvatures ${}_{(1)}b, {}_{(2)}b, \cdots, {}_{(n)}b$. The *gauge function* ϕ has the transformation law

$$\bar{\phi} = \phi - \sigma. \tag{2.6}$$

The *gauge vector* η^a given by

$$\eta^a = \phi_{,i}g^{ij}y_{,j}{}^a + \mu^a \tag{2.7}$$

has the transformation law

$$\bar{\eta}_a = \eta_a - \sigma_{,a} \tag{2.8}$$

relative to conformal maps of R_{n+1}. The *conformal measure tensors* $A_{\alpha\beta}, G_{ij}$ and the *conformal coefficients of connection* $\Gamma_{\beta\gamma}{}^{\alpha}, \Gamma_{jk}{}^{i}$ are defined by

$$A_{\alpha\beta} = e^{2\phi}a_{\alpha\beta}, \qquad G_{ij} = e^{2\phi}g_{ij} \tag{2.9}$$

$$\Gamma_{\beta\gamma}{}^{\alpha} = \left\{ \begin{matrix} \alpha \\ \beta\gamma \end{matrix} \right\} + \delta_{\beta}{}^{\alpha}\eta_{\gamma} + \delta_{\gamma}{}^{\alpha}\eta_{\beta} - a_{\beta\gamma}a^{\alpha\delta}\eta_{\delta}$$

(2.10)

$$\Gamma_{jk}{}^{i} = \left\{ \begin{matrix} i \\ jk \end{matrix} \right\} + \delta_{j}{}^{i}\phi_{,k} + \delta_{k}{}^{i}\phi_{,j} - g_{jk}g^{ih}\phi_{,h}$$

and remain unchanged by conformal transformations of R_{n+1}. The $\Gamma_{jk}{}^{i}$ are the Christoffel symbols of the second kind formed with respect to G_{ij}. The δ's in (2.10) are the Kronecker deltas.

Conformal length or *conformal distance* S is measured by means of $A_{\alpha\beta}$, G_{ij}; that is, for a displacement dy^α from a point of V_n,

$$dS^2 = A_{\alpha\beta}dy^\alpha dy^\beta. \tag{2.11}$$

If dy^α lies in the tangent vector space of V_n and its surface components are dx^i,

$$dS^2 = G_{ij}dx^i dx^j. \tag{2.12}$$

It is clear that

$$dS = e^\phi ds \tag{2.13}$$

where ds is the element of distance of R_{n+1} defined by $ds^2 = a_{\alpha\beta}dy^\alpha dy^\beta$ or on V_n, by $ds^2 = g_{ij}dx^i dx^j$.

Conformal differentiation, denoted by a colon, is defined as covariant tensor differentiation relative to the conformal coefficients of connection. For example,

$$T^{\alpha i}_{\beta j:k} = \partial T_{\beta j}{}^{\alpha i}/\partial x^k + T_{\beta j}{}^{\gamma i}\Gamma_{\gamma\delta}{}^{\alpha}y_{,k}{}^{\delta} + T_{\beta j}{}^{\alpha h}\Gamma_{hk}{}^{i} - T_{\gamma j}{}^{\alpha i}\Gamma_{\beta\delta}{}^{\gamma}y_{,k}{}^{\delta} - T_{\beta h}{}^{\alpha i}\Gamma_{jk}{}^{h}. \tag{2.14}$$

The conformal derivative of a conformal tensor is a conformal tensor. Other tensors defined on V_n which are invariant relative to conformal mappings of R_{n+1} are:

Conformal normal measure tensor B_{ij}

$$B_{ij} = e^\phi\ [b_{ij} - \mu g_{ij}]. \tag{2.15}$$

Deviation tensor E_{ij}

$$\begin{aligned} E_{ij} &= \mu_\alpha G_{ij}{}^{\alpha} - \phi_{,ij} + \phi_{,i}\phi_{,j} + \tfrac{1}{2}[\mu^2 - g^{hk}\phi_{,h}\phi_{,k}]g_{ij} \\ &= \mu_\alpha G_{ij}{}^{\alpha} - \phi_{:ij} - \phi_{:i}\phi_{:j} + \tfrac{1}{2}[e^{-2\phi}\mu^2 + G^{hk}\phi_{:h}\phi_{:k}]G_{ij}. \end{aligned} \tag{2.16}$$

Deviation vector F_i

$$F_i = (1/(n-1))\ldots \tag{2.17}$$

(*Surface*) *conformal Riemann tensor* $K_{jkl}{}^{i}$

$$K_{jkl}{}^{i} = (\partial/\partial x^k)T_{jl}{}^{i} - (\partial/\partial x^l)\Gamma_{jk}{}^{i} + \Gamma_{jl}{}^{h}\Gamma_{hk}{}^{i} - \Gamma_{jk}{}^{h}\Gamma_{hl}{}^{i}. \tag{2.18}$$

The tensors G_{ij}, B_{ij}, E_{ij} are called the *first, second and third conformal fundamental tensors* of V_n, respectively, and serve as the coefficients of the *conformal fundamental forms* of V_n. The tensor B_{ij} satisfies the algebraic relations

$$B_{ij}G^{ij} = 0, \qquad B_{ij}B_{kl}G^{ik}G^{jl} = n. \tag{2.19}$$

Differential conditions on the conformal tensors are

$$B_{ik}B_{hj} - B_{ij}B_{hk} = K_{hijk} + G_{ij}E_{hk} - G_{ik}E_{hj} + G_{hk}E_{ij} - G_{hj}E_{ik} \tag{2.20}$$

$$B_{ij:k} - B_{ik:j} = G_{ik}F_j - G_{ij}F_k \tag{2.21}$$

$$F_{j:k} - K_{k:j} = B_{hk}E_j^h - B_{hj}E_k^h \tag{2.22}$$

$$E_{ij:k} - E_{ik:j} = B_{ij}F_k - B_{ik}F_j. \tag{2.23}$$

The basic theorems are:

(1) *Equivalence theorem*: If the conformal fundamental forms of two hypersurfaces V_n and $\bar{V}_n$ are equal, then a conformal transformation of R_{n+1} on itself exists so that V_n is mapped on $\bar{V}_n$.

(2) *Existence theorem*: A hypersurface in R_{n+1} always exists having prescribed conformal fundamental forms, provided the tensor coefficients of these forms satisfy equations (2.19) to (2.23).

3. The case $n=2$. In surface theory $n=2$, $m=3$. In this case, we may choose the lines of curvature of V_2 as parametric. We write the surface parameters x^1, x^2 as u, v, respectively, and the components of the conformal measure tensor as

$$\begin{array}{lll} G_{11}=X & G_{12}=0 & G_{22}=Y \\ G^{11}=X^{-1} & G^{12}=0 & G^{22}=Y^{-1}. \end{array} \tag{3.1}$$

According to (2.15), the principal directions determined by B_{ij} relative to G_{ij} are the lines of curvature. (This fact establishes the conformal invariance of the lines of curvature.) It follows from (2.19) that the components of B_{ij} are

$$B_{11}=X \qquad B_{12}=0 \qquad B_{22}=-Y. \tag{3.2}$$

The Christoffel symbols Γ_{jk}^{i} defined by (2.10) become

$$\begin{array}{lll} \Gamma_{11}^{1}= X_u/2X & \Gamma_{12}^{1}=X_v/2X & \Gamma_{22}^{1}=-Y_u/2X \\ \Gamma_{11}^{2}=-X_v/2Y & \Gamma_{12}^{2}=Y_u/2Y & \Gamma_{22}^{2}= Y_v/2Y. \end{array} \tag{3.3}$$

From (3.2) and (3.3),

$$\begin{array}{llll} B_{11:1}=B_{11:2}= & B_{22:1}=B_{22:2}=0 \\ B_{12:1}=-X_v & B_{12:2}=Y_u. \end{array} \tag{3.4}$$

Also, from (2.17), (3.1), (3.3) and (3.4),

$$(3.5)\qquad \begin{aligned} F_1 &= [\log Y]_u & F_2 &= -[\log X]_v \\ F_{1:2} &= [\log Y]_{uv} & F_{2:1} &= -[\log X]_{uv}. \end{aligned}$$

Using the above equations, it is easy to show that equations (2.21) are true identically when $n=2$. Equations (2.20) are equivalent to

$$(3.6)\qquad K = 2 - 2E$$

where K and E are defined by

$$(3.7)\qquad K = K_{jki}{}^{i}G^{jk} = 1/2(XY)^2 \times [2XY(X_{vv}+Y_{uu}) - Y_u(XY)_u - X_v(XY)_v]$$

$$(3.8)\qquad E = E_{ij}G^{ij}.$$

Since the *conformal Gaussian curvature* J of V_2 is defined by

$$(3.9)\qquad J = K_{1212}/XY$$

it follows from (3.7) and (3.9) that

$$(3.10)\qquad K = -2J$$

so that (3.6) may be written as

$$(3.11)\qquad E = J + 1.$$

Equation (2.22) becomes

$$(3.12)\qquad [\log \sqrt{XY}]_{uv} + E_{12} = 0$$

and (2.23) becomes

$$(3.13)\qquad \partial E_{11}/\partial v = \partial E_{12}/\partial u + \tfrac{1}{2}E_{12}\,[\log Y/X]_u + \tfrac{1}{2}X_v(E-2)$$

$$(3.14)\qquad \partial E_{22}/\partial v = \partial E_{12}/\partial v - \tfrac{1}{2}E_{12}\,[\log Y/X]_v + \tfrac{1}{2}Y_u(E-2).$$

If ${}_{(1)}L^i$, ${}_{(2)}L^i$ are the unit conformal tangent vectors to the lines of curvature, then, according to (3.1) and (3.2), G_{ij} and B_{ij} have the canonical representation

$$(3.15)\qquad \begin{aligned} G_{ij} &= {}_{(1)}L_i\,{}_{(1)}L_j + {}_{(2)}L_i\,{}_{(2)}L_j \\ B_{ij} &= {}_{(1)}L_i\,{}_{(1)}L_j - {}_{(2)}L_i\,{}_{(2)}L_j. \end{aligned}$$

It follows from the known theory [1, pp. 107-113] that any symmetric second order tensor T_{ij} whose principal directions are tangent to the lines of curvature must have the canonical representation

$$(3.16)\qquad T_{ij} = {}_{(1)}a\,{}_{(1)}L_i\,{}_{(1)}L_j + {}_{(2)}a\,{}_{(2)}L_i\,{}_{(2)}L_j$$

where ${}_{(1)}a$ and ${}_{(2)}a$ are the principal invariants of T_{ij}. In virtue of (3.15), (3.16) implies

$$T_{ij} = \tfrac{1}{2}[({}_{(1)}a + {}_{(2)}a)G_{ij} + ({}_{(1)}a - {}_{(2)}a)B_{ij}] \tag{3.17}$$

and conversely. We also note that

$$\mu = \tfrac{1}{2}[{}_{(1)}b + {}_{(2)}b], \qquad e^{\phi} = \tfrac{1}{2}\,|\,{}_{(1)}b - {}_{(2)}b\,|. \tag{3.18}$$

In view of our choice of parameters u, v,

$${}_{(1)}L_i = X^{\frac{1}{2}}\delta_i^{\,1} \qquad {}_{(2)}L_i = Y^{\frac{1}{2}}\delta_i^{\,2}$$

where the δ's are the Kronecker deltas. The *conformal (surface) principal normal vectors* ${}_{(1)}N_i$, ${}_{(2)}N_i$ of the lines of curvature are defined as

$$\begin{aligned} {}_{(1)}N_i &= d\,{}_{(1)}L_i/d\,{}_{(1)}S - \Gamma_{ij}^{h}\,{}_{(1)}L_h\,{}_{(1)}L^j = -\,{}_{(1)}N\cdot{}_{(2)}L_i \\ {}_{(2)}N_i &= d\,{}_{(2)}L_i/d\,{}_{(2)}S - \Gamma_{ij}^{h}\,{}_{(2)}L_h\,{}_{(2)}L^j = -\,{}_{(2)}N\,{}_{(1)}L_i \end{aligned} \tag{3.19}$$

where $d\,{}_{(1)}S$ and $d\,{}_{(2)}S$ indicates directional differentiation with respect to S in the directions ${}_{(1)}L_i$ and ${}_{(2)}L_i$, respectively, and ${}_{(1)}N$ and ${}_{(2)}N$ are the corresponding *conformal (surface) curvatures* of the lines of curvature. ${}_{(1)}N$ and ${}_{(2)}N$ are the analogues of "geodesic curvature." From (3.19), we obtain

$${}_{(1)}N = X_v/2XY^{\frac{1}{2}} \qquad {}_{(2)}N = Y_u/2YX^{\frac{1}{2}}. \tag{3.20}$$

From (3.20), it follows that

$$\begin{aligned} d\,{}_{(1)}N/d\,{}_{(1)}S &= 1/2\sqrt{XY}\{[\log X]_{uv} - \tfrac{1}{2}[\log X]_v[\log Y]_u\} \\ d\,{}_{(2)}N/d\,{}_{(2)}S &= 1/2\sqrt{XY}\{[\log Y]_{uv} - \tfrac{1}{2}[\log X]_v[\log Y]_u\}. \end{aligned} \tag{3.21}$$

4. Gradient surfaces. The *deviation* of V_2 for the direction determined by the unit conformal vector L^i is $E_{ij}L^iL^j$. Since E_{ij} is a symmetric tensor, it follows from the known theory that its principal directions relative to G_{ij} are two orthogonal directions at each point of V_2. These are the *deviation directions* and the corresponding curves are the *lines of deviation*. The corresponding deviations are the *principal deviations* ${}_{(1)}E$ and ${}_{(2)}E$. They are the roots of the determinantal equation $|\,E_{ij} - \rho G_{ij}\,| = 0$. It follows from (3.8) $F_{;ij}\,{}_{(2)}L^i\,{}_{(1)}L^j = 0$.

$$E = {}_{(1)}E + {}_{(2)}E. \tag{4.1}$$

Of course, the deviation directions are conformally invariant. When do the lines of curvature and the lines of deviation of V_2 coincide? Since the lines of curvature are parametric, the necessary and sufficient condition is $E_{12} = 0$. It follows from (3.12) that this condition is equivalent to

(4.2) $$[\log \sqrt{XY}]_{uv} = 0$$

which leads to

$$X/f^2(u) = g^2(v)/Y = Q(u, v).$$

Consequently, after a parameter transformation

(4.3) $$u' = \int f(u)\,du, \qquad v' = \int g(v)\,dv$$

which preserves the parametric directions,

$$dS^2 = Q \cdot du'^2 + 1/Q \cdot dv'^2.$$

Dropping the primes, we may suppose

(4.4) $$X = Q \qquad Y = 1/Q.$$

Conversely, (4.4) implies (4.2) and hence $E_{12} = 0$. Now the conformal area of any curvilinear rectangle bounded by the lines of curvature of parameters u_1, u_2, v_1, v_2 is

$$\int\int \sqrt{XY}\,du\,dv = (u_2 - u_1)(v_2 - v_1).$$

Hence the lines of curvature form a (conformal) equiareal pattern. This proves the following theorem:

THEOREM 4.1. *A necessary and sufficient condition that the lines of curvature and the lines of deviation of a surface coincide is that the lines of curvature be an equiareal net (relative to the element of conformal area).*

According to (3.5), (4.2) is also equivalent to

(4.5) $$F_{1:2} = F_{2:1}.$$

As is well known, (4.5) is the condition that F_i be a gradient, that is,

(4.6) $$F_i = F_{:i}$$

where F is a scalar. It follows from (3.5) that F may be defined as $\log Y$ in this case. We call a V_2 for which (4.6) is true a *gradient surface*. The preceding discussion proves

THEOREM 4.2. *The lines of curvature and the lines of deviation of a V_2 coincide if and only if the V_2 is a gradient surface.*

Finally, we note that in accordance with (3.17), if V_2 is a gradient surface,

$$E_{ij} = \tfrac{1}{2}[EG_{ij} + E'B_{ij}] \tag{4.7}$$

where

$$E' = {}_{(1)}E - {}_{(2)}E. \tag{4.8}$$

5. Isothermic surfaces. In classical differential geometry, surfaces whose lines of curvature form an isothermal net are called isothermic. The isothermal character of a net is conformally invariant since such a net is the conformal image of a rectangular net of straight lines. Hence isothermic surfaces remain isothermic under conformal transformations of R_3. We derive a conformal characterization of these surfaces below.

If V_2 is isothermic, it follows from (2.9) and (3.1) that

$$X = Y \tag{5.1}$$

in some parameter system for which the lines of curvature are parametric, and conversely. It follows from (3.5) that

$$F_{1:2} + F_{2:1} = 0. \tag{5.2}$$

Conversely, (5.2) implies (5.1) after parameter transformations of the type (4.3). Equation (5.2) implies that the principal directions of the symmetric tensor $F_{i:j} + F_{j:i}$ relative to G_{ij} are tangent to the lines of curvature. The concluding remarks of **3** leading to (3.17) then prove

THEOREM 5.1. *The tensor $F_{i:j} + F_{j:i}$ is a linear combination of the first and second conformal fundamental tensors of V_2 if and only if V_2 is isothermic.*

We also note that

$$F_{1:2} + F_{2:1} = 2\sqrt{XY}[d_{(2)}N/d_{(2)}S - d_{(1)}N/d_{(1)}S]$$

as follows from (3.5) and (3.21). As a consequence of the preceding theorem, *the equation*

$$d_{(1)}N/d_{(1)}S = d_{(2)}N/d_{(2)}S$$

is a characteristic property of isothermic surfaces.

6. Surfaces of constant conformal Gaussian curvature. A surface for which J is constant is called a *surface of constant conformal Gaussian curvature*. Exactly as in classical differential geometry, such surfaces admit three-parameter groups of point transformations on themselves (*conformal motions*)

which preserve the element of conformal length dS. They also enjoy those other properties of surfaces of constant curvature which are derivable from the first fundamental form of the surface. A different kind of property of these surfaces follows directly from (3.11) and (4.1): *The sum of the principal deviations of a V_2 is constant if and only if the V_2 has constant conformal Gaussian curvature.*

The simplest of these surfaces are those for which J is zero and whose conformal metric, referred to the lines of curvature, is Cartesian. In this case

$$X = Y = 1 \tag{6.1}$$

so that

$$\begin{aligned} G_{ij}dx^i dx^j &= du^2 + dv^2 \\ B_{ij}dx^i dx^j &= du^2 - dv^2. \end{aligned} \tag{6.2}$$

It follows from (3.12) that $E_{12} = 0$ and from (3.13) and (3.14) that

$$E_{11} = E_{11}(u), \qquad E_{22} = E_{22}(v). \tag{6.3}$$

Since (4.1) is true, it follows from (3.11) and (6.3) that

$$E_{11} = c, \qquad E_{12} = 0, \qquad E_{22} = 1 - c \tag{6.4}$$

where c is a constant. Also, according to (3.5) and (6.1),

$$F_1 = 0, \qquad F_2 = 0. \tag{6.5}$$

Conversely, (6.5) implies (6.1) in view of (3.5). Now it is shown in a forthcoming paper that the line of curvature, $v =$ constant, is a circle if and only if $F_1 = 0$ along its length. Hence (6.5) implies that all the lines of curvature of V_2 are circular. Of course, this is a conformally invariant property. It is a characteristic property of Dupin cyclides and proves

THEOREM 6.1. *If the lines of curvature of a V_2 form a Euclidean Cartesian net relative to the conformal element of length on V_2, then V_2 is a Dupin cyclide, and conversely.*

The simplest Dupin cyclides are the right circular cone, right circular cylinder and proper torus. The first and second conformal fundamental forms of each of these surfaces are given by (6.2) after a suitable choice of parameters. Below, we determine the third conformal fundamental form in each case, using the definitions of **2** and **3**. The equations and components of the fundamental metric tensors of a right circular cone are

$$\begin{array}{lll} y^1 = \omega\cos\alpha & y^2 = \omega\sin\alpha & y^3 = a\omega \\ g_{11} = \omega^2 & g_{12} = 0 & g_{22} = 1 + a^2 \\ b_{11} = 0 & b_{12} = 0 & b_{22} = a\omega/(\sqrt{1+a^2}). \end{array}$$

After the parameter transformation

$$u = a\alpha/(2\sqrt{1+a^2}) \qquad v = (a/2)\log\omega$$

we find that (6.2) is true and that

(6.6) $$E_{ij}dx^i dx^j = ((3/2) + 2/a^2)du^2 + (-(1/2) - (2/a^2))dv^2.$$

The corresponding equations and tensor components for a right circular cylinder are

$$\begin{array}{lll} y^1 = a\cos\alpha & y^2 = a\sin\alpha & y^3 = \omega \\ g_{11} = a^2 & g_{12} = 0 & g_{22} = 1 \\ b_{11} = 0 & b_{12} = 0 & b_{22} = a. \end{array}$$

After the parameter transformation

$$u = 1\alpha/2 \qquad v = 1\omega/2a$$

we obtain (6.2) and

(6.7) $$E_{ij}dx^i dx^j = (3du^2)/2 - (1dv^2)/2.$$

The corresponding equations and tensor components for a proper torus are

$$\begin{array}{lll} y^1 = (a + d\cos\beta)\cos\alpha & y^2 = (a + d\cos\beta)\sin\alpha & y^3 = d\sin\beta, \quad a > d \\ g_{11} = (a + d\cos\beta)^2 & g_{12} = 0 & g_{22} = d^2 \\ b_{11} = (a + d\cos\beta)\cos\beta & b_{12} = 0 & b_{22} = d. \end{array}$$

After the parameter transformation

$$u = a\alpha/2d \qquad v = (a/2)\int[d\beta/(a + d\cos\beta)]$$

we establish (6.2) and

(6.8) $$E_{ij}dx^i dx^j = (-(1/2) + 2d^2/a^2)du^2 + ((3/2) - (2d^2/a^2))dv^2.$$

In all the above equations, a, d are constants and α, β, ω are variables.

A comparison of these results with (6.3) shows that every possible form $E_{ij}dx^i dx^j$ for a Dupin cyclic is actually realized by the cone, cylinder and proper [2] torus. Now all Dupin cyclides have the first and second conformal fundamental forms which appear in (6.2). Therefore, in accordance with the conformal equivalence theorem noted in 2, we have the following theorem:

THEOREM 6.2. *Every Dupin cyclide is conformally equivalent to a right circular cone, right circular cylinder or proper torus, and conversely.*

---

[2] The improper torus for which $a = d$ or $a < d$ is conformally equivalent to the cylinder or cone, respectively, as may be seen by a comparison of $E_{ij}dx^i dx^j$ for these surfaces.

As an example of a surface having non-zero constant conformal Gaussian curvature we mention the catenoid whose equations are

$$y^1 = \omega \cos \alpha \qquad y^2 = \omega \sin \alpha \qquad y^3 = a \log (\omega + \sqrt{\omega^2 - a^2}).$$

From these equations we obtain

$$\begin{array}{lll} g_{11} = \omega^2 & g_{12} = 0 & g_{22} = \omega^2/(\omega^2 - a^2) \\ b_{11} = a & b_{12} = 0 & b_{22} = a/(a^2 - \omega^2). \end{array}$$

After a straightforward calculation and the parameter transformation

$$u = \alpha \qquad v = \text{arc sec } (\omega/a)$$

we find that

$$\begin{aligned} G_{ij}dx^i dx^j &= \cos^2 v\, du^2 + dv^2 \\ B_{ij}dx^i dx^j &= \cos^2 v\, du^2 - dv^2 \\ E_{ij}dx^i dx^j &= \qquad\qquad 2dv^2. \end{aligned}$$

From the last of these equations and (3.11) and (4.1), it follows that $J = 1$.

We note that, as is stated earlier, if

$$F_1 = 0 \tag{6.9}$$

throughout a V_2, then all the lines of curvature of V_2, $v =$ constant, are circles. Hence (6.9) is a characteristic property of any surface which is the envelope of an arbitrary one parameter family of spheres (if the lines of curvature are parametric). In any parameter system the corresponding condition is

$$F_{i\,(1)}L^i = 0. \tag{6.10}$$

Since, from (3.15),

$${}_{(1)}L_{i\,(1)}L_j = \tfrac{1}{2}[G_{ij} + B_{ij}]$$

(6.10) may be written as

$$F_i + B_{ij}F^j = 0$$

where $F^j = G^{jk}F_k$.

7. Isothermic gradient surfaces. According to (4.5) and (5.2), isothermic gradient surfaces are characterized by the equations

$$F_{1:2} = F_{2:1} = 0. \tag{7.1}$$

Since $F_i = F_{:i}$ in this case, we may write the above equation in the form

$$F_{:ij\,(1)}L^i{}_{(2)}L^j = F_{:ij\,(2)}L^i{}_{(1)}L^j = 0.$$

This means that the principal directions of $F_{i:j}$ are tangent to the lines of curvature so that *$F_{i:j}$ is a linear combination of G_{ij} and B_{ij}.*

We now determine the distinctive character of the conformal fundamental forms of such surfaces. Since V_2 is a gradient surface, (4.4) is true. The isothermic character of V_2 implies (5.2) which, using (3.5) and (4.4), leads to

$$Q = Q_2(v)/Q_1(u) \tag{7.2}$$

where Q_1 and Q_2 are, respectively, functions of u and v only. Conversely (7.2) implies (7.1) which characterizes isothermic surfaces.

It remains to impose the condition that

$$dS^2 = (Q_2/Q_1)du^2 + (Q_1/Q_2)dv^2 \tag{7.3}$$

be the first conformal fundamental form of an actual surface. In view of the existence theorem stated in **2**, it is necessary and sufficient to impose conditions (3.11) to (3.14). Equation (3.12) is satisfied identically since (4.4), (7.2) and the equation $E_{12} = 0$ are true. According to (3.7), (3.10), (4.4) and (7.2),

$$J = -\tfrac{1}{2}[Q_2''/Q_1 + Q_1''/Q_2] \tag{7.4}$$

where the primes refer to differentiation, with respect to the independent variable, of the functions Q_1, Q_2. From (3.11), (4.1) and (4.8), we obtain

$$2\,{}_{(1)}E = 1 + J + E', \qquad 2\,{}_{(2)}E = 1 + J - E'. \tag{7.5}$$

Also, since

$$E_{11} = {}_{(1)}E(Q_2/Q_1), \qquad E_{22} = {}_{(2)}E(Q_1/Q_2), \qquad E_{12} = 0$$

(3.13) and (3.14) become

$${}_{(1)}E_v = -(Q'_2/Q_1)(E'/2 + 1), \qquad {}_{(2)}E_u = (Q'_1/Q_2)(E'/2 - 1). \tag{7.6}$$

Differentiation of (7.5) yields

$$\begin{aligned} {}_{(1)}E_v &= E_v'/2 - \tfrac{1}{4}(Q_2'''/Q_1 - Q_2'Q_1''/Q_2^2) \\ {}_{(2)}E_u &= -E_u'/2 - \tfrac{1}{4}(Q_1'''/Q_2 - Q_1'Q_2''/Q_1^2) \end{aligned} \tag{7.7}$$

after use is made of (7.4). From (7.6) and (7.7), we obtain

$$\begin{aligned} E_v' &= -(Q'_2/Q_2)(2 + E') + \tfrac{1}{2}(Q_2'''/Q_1 - Q_2'Q_1''/Q_2^2) \\ E_u' &= \quad (Q'_1/Q_1)(2 - E') - \tfrac{1}{2}(Q_1'''/Q_2 - Q'Q_2''/Q_1^2). \end{aligned}$$

The integrability condition of these equations is $E'_{vu} = E'_{uv}$ or

$$\begin{aligned} 8Q_1'Q_2'/Q_1Q_2 + Q_1'Q_2'''/Q_1^2 + Q_2'Q_1'''/Q_2^2 \\ + Q_1'Q_1''Q_2'/Q_1Q_2^2 + Q_2'Q_2''Q_1'/Q_2Q_1^2 = 0. \end{aligned} \tag{7.8}$$

Let us suppose that $Q_1'Q_2' \neq 0$. Then multiplication of (7.8) by $Q_1^2Q_2^2/Q_1'Q_2'$ and successive differentiation with respect to u and then v leads to the equation, $Q_1'Q_2' = 0$, which contradicts the hypothesis. The remaining possibility, $Q_1'Q_2' = 0$, is clearly a solution of (7.8). Without loss of generality (interchanging u and v if necessary) we may assume that

$$Q_1' = 0, \qquad Q_1 = 1, \qquad Q_2 = Q(v).$$

In this case, the form (7.3) may be written as

$$dS^2 = Q^2(v)\,du^2 + dv^2 \tag{7.9}$$

after the parameter transformation

$$u' = u, \qquad v' = \int (dv/\sqrt{Q})$$

and writing $Q^2(v')$ for $Q[v(v')]$ and dropping primes.

We have shown that the first conformal fundamental form of every isothermic gradient surface may be written as (7.9). Conversely, if $Q(v)$ in (7.9) is an arbitrary function of v, the surface is isothermic gradient. Indeed, from (3.11) to (3.14), we find that the only E_{ij} which are possible are the coefficients of the form

$$\begin{aligned} E_{ij}dx^i dx^j = {} & [-(1/2)Q^2 - (1/2)Q'^2 + c]\,du^2 \\ & + [3/2 + \frac{1}{2}\frac{Q'^2}{Q^2} - Q''/Q - c/Q^2]\,dv^2 \end{aligned} \tag{7.10}$$

where c is a constant. Then the existence theorem of **2** affirms the existence of surfaces for which (7.9) and (7.10) are the first and third conformal fundamental forms. Of course, the second conformal fundamental form is

$$B_{ij}dx^i dx^j = Q^2(v)\,du^2 - dv^2. \tag{7.11}$$

It is clear that we may also choose u, v so that the lines of curvature are parametric and

$$dS^2 = Q^2(v)\,[du^2 + dv^2]. \tag{7.12}$$

In this case, the second and third conformal fundamental forms are

$$\begin{aligned} B_{ij}dx^i dx^j &= Q^2(v)\,[du^2 - dv^2] \\ E_{ij}dx^i dx^j &= [-(1/2)Q^2 - \frac{1}{2}\frac{Q'^2}{Q^2} + c]\,du^2 \\ &\quad + [(3/2)Q^2 + \frac{3}{2}\frac{Q'^2}{Q^2} - Q''/Q - c]\,dv^2 \end{aligned} \tag{7.13}$$

where c is the same constant that occurs in (7.10). We also note, as follows

from (3.5) and (7.12), that (6.9) is true. Hence all the lines of curvature, v = constant, are circles. We state some of these facts in the following theorem:

THEOREM 7.1. *If V_2 is an isothermic gradient surface, its first conformal fundamental form is defined by* (7.9) *in some parameter system for which the lines of curvature are parametric, and conversely. The other conformal fundamental forms are then defined by* (7.10) *and* (7.11).

The simplest isothermic gradient surfaces are those for which $Q(v)$ in (7.12) is a constant. In this case, in accordance with Theorem 6.1, the surface is a Dupin cyclide. We now consider the problem of characterizing all isothermic gradient surfaces in an analogous manner.

Let V_2 be an isothermic gradient surface and suppose that its (metric) first and second fundamental forms referred to the parameters u, v of (7.12) are

$$\begin{aligned} g_{ij}dx^i dx^j &= e^{2\alpha}(du^2 + dv^2) \\ b_{ij}dx^i dx^j &= e^{\alpha}(\gamma du^2 + \delta dv^2) \end{aligned} \tag{7.14}$$

where α, γ, δ are functions of u and v. In these parameters, the classical Gauss-Codazzi equations become

$$\gamma\delta + \alpha_{uu} + \alpha_{vv} = 0 \tag{7.15}$$

$$\gamma_v - \delta\alpha_v = 0 \tag{7.16}$$

$$\delta_u - \gamma\alpha_u = 0. \tag{7.17}$$

Now, according to (3.18),

$$e^{2\phi} = \tfrac{1}{4}(e^{-\alpha}\gamma - e^{-\alpha}\delta)^2$$

so that, from (2.13),

$$dS^2 = \tfrac{1}{4}(\gamma - \delta)^2(du^2 + dv^2). \tag{7.18}$$

A comparison of (7.12) and (7.18) shows that

$$\gamma = \delta + 2Q. \tag{7.19}$$

Then integration of (7.17) leads to

$$\delta + 2Q = e^{\alpha}f(v) \tag{7.20}$$

where $f(v)$ is an arbitrary function of v. Substitution of (7.19) and (7.20) in (7.16) and integration results in the equation [3]

[3] Since the geodesic curvatures of the lines of curvature of a V_2 satisfying (7.21), v = constant and u = constant, are $-h'(v)$ and $-g'(u)$, respectively, each line of curvature has constant geodesic curvature.

(7.21) $$e^{-a} = g(u) + h(v)$$

where

(7.22) $$h'(v) = f'(v)/2Q.$$

If (7.19), (7.20) and (7.21) are substituted in (7.15), we obtain

(7.23) $$[f^2 - 2Qfh + h'^2 - hh'']$$
$$+ [g'^2 - gg''] - g[2fQ + h''] - hg'' = 0.$$

If (7.23) is differentiated successively with respect to u and v, we obtain

(7.24) $$g'[(2fQ)' + h'''] + h'g''' = 0.$$

(I) Suppose that $g'h' \neq 0$. Then division of (7.24) by $g'h'$ leads to two separate equations which when integrated yield the equations

(7.25) $$g'' = c_1 g + c_2$$

(7.26) $$2fQ + h'' = -c_1 h + c_3$$

where c_1, c_2, c_3 are constants. We distinguish several cases:

(I_1) $c_1 = 0$. The solution of (7.25) is

(7.27) $$g = \tfrac{1}{2}c_2 u^2 + c_4$$

after a translation of u, if necessary. Substitution of this result and (7.26) in (7.23) leads to

(7.28) $$c_1 = c_3$$

(7.29) $$f^2 - 2fhQ + h'^2 - hh'' - c_2 h - 2c_2 c_4 = 0.$$

Use of (7.26) permits us to write (7.29) as

(7.30) $$f^2 + h'^2 - 2c_2 h - 2c_2 c_4 = 0.$$

According to (7.22), if $f' = 0$ then $h' = 0$. Since this case has been excluded, f' is not zero. Then (7.26) is a consequence of (7.22) and (7.30) which are the final conditions on $f(v)$ and $h(v)$ in this case.

(I_2) $c_1 = -b^2$. The solution of (7.25) is

(7.31) $$g = c_4 \cos bu + c_2/b^2$$

after a translation of u, if necessary. Substitution of this result and (7.26) in (7.23) yields (7.28) and

(7.32) $$f^2 - 2fhQ + h'^2 - hh'' + c_4^2 b^2 - c_2 h - c_2^2/b^2 = 0.$$

By means of (7.26), (7.32) becomes

(7.33) $$f^2 + h'^2 - b^2 h^2 + c_4^2 b^2 - 2c_2 h - c_2^2/b^2 = 0.$$

Since f' differs from zero, (7.26) is a consequence of (7.22) and (7.33). These are the final two conditions on $f(v)$ and $h(v)$ in this case.

(I_3) $c_1 = +b^2$. The solution of (7.25) is

$$\begin{aligned} &\text{(i)} \quad g = c_4 \cosh bu - c_2/b^2 \\ &\text{(ii)} \quad g = c_4 \sinh bu - c_2/b^2 \\ &\text{(iii)} \quad g = c_4 e^{bu} - c_2/b^2 \end{aligned} \tag{7.34}$$

after a translation of u, if necessary. If (7.34) is substituted in (7.23) and use is made of (7.26), we find that (7.28) is true and also obtain

$$f^2 - 2fhQ + h'^2 - hh'' - \epsilon c_4{}^2 b^2 - c_2 h + c_2{}^2/b^2 = 0 \tag{7.35}$$

where $\epsilon = +1, -1, 0$ in cases (i), (ii) and (iii), respectively. Using (7.26), (7.35) becomes

$$f^2 + h'^2 + b^2 h^2 - \epsilon c_4{}^2 b^2 - 2c_2 h + c_2{}^2/b^2 = 0. \tag{7.36}$$

As in the previous two cases, (7.22) and (7.36) are the final conditions on $f(u)$ and $h(v)$.

(II) We now assume that $g'h' = 0$. Then (7.24) is true. We consider several cases:

(II_1) $g' = h' = 0$. We may assume, without loss of generality, that

$$g = 1, \qquad h = 0. \tag{7.37}$$

Substitution of (7.37) in (7.23) leads to

$$f = 0 \tag{7.38}$$

which is also consistent with (7.22). Hence (7.37) and (7.38) are the final conditions in this case.

(II_2) $g' = 0$, $h' \neq 0$. Hence

$$g = c_4. \tag{7.39}$$

Then (7.23) becomes, after use of (7.22),

$$f^2 + h'^2 = b^2 (c_4 + h)^2. \tag{7.40}$$

Equations (7.22) and (7.40) are the final conditions.

(II_3) $h' = 0$, $g' \neq 0$. Taking account of (7.22),

$$h = c_4, \qquad f = c_5, \tag{7.41}$$

so that (7.23) becomes

$$(7.42) \qquad c_5^2 - 2c_4c_5Q + g'^2 - gg'' - 2c_5gQ - c_4g'' = 0.$$

Successive differentiation with respect to u and v shows that $c_5 = 0$ since $Q' \neq 0$. Hence the solution of (7.42) is

$$(7.45) \qquad g = c_6e^{bu} - c_4.$$

The solutions obtained under the above six categories exhaust the possibilities for a given function $Q(v)$. In all cases, the first and second conformal fundamental forms are the same but the third conformal fundamental form (7.13) may differ in the additive constant c. We shall calculate E_{11} for each of the above solutions in order to distinguish the various conformally different types. From (7.19), (7.20) and (7.21), the principal normal curvatures of any of these surfaces are

$$(7.46) \qquad {}_{(1)}b = f(v), \qquad {}_{(2)}b = f(v) - 2Q[g(u) + h(v)].$$

According to (3.18),

$$(7.47) \qquad e^{\phi} = Q(g + h), \qquad \mu = f - Q(g + h).$$

Also, from (7.12),

$$\Gamma_{11}{}^1 = 0 \qquad \Gamma_{11}{}^2 = -Q'/Q.$$

From these equations, we find that

$$\phi_{:1} = g'/(g + h) \qquad \phi_{:2} = Q'/Q + h'/(g + h)$$
$$\phi_{:11} = [g''(g + h) - g'^2]/(g + h)^2 + Q'^2/Q^2 + 2h'Q'/(g + h).$$

From the second set of equations (2.14) and the above equations,

$$(7.48) \qquad E_{11} = -Q^2/2 - \frac{1}{2}\frac{Q'^2}{Q^2} + \tfrac{1}{2}Q^2(e^{-\phi}\mu + 1)^2 + [1/(g + h)^2][\tfrac{1}{2}g'^2 - g''(g + h) + \tfrac{1}{2}h'^2].$$

But, from (7.47),

$$e^{-\phi}\mu + 1 = f/Q(g + h).$$

Substitution of this result in (7.48) gives

$$(7.49) \qquad E_{11} = -Q^2/2 - \frac{1}{2}\frac{Q'^2}{Q^2} + 1/2(g + h)^2 \times [g'^2 - 2gg'' - 2hg'' + h'^2 + f^2].$$

Comparison of (7.13) and (7.49) shows that

$$c = [1/2(g + h)^2][g'^2 - 2gg'' - 2hg'' + h'^2 + f^2].$$

If the right member of this equation be evaluated by means of the conditions on $f(v), h(v)$ in each case, we find that

$$(7.50)\qquad \begin{array}{lll} \text{(i)} & c = 0 & \text{in cases } I_1, II_1 \\ \text{(ii)} & c = \tfrac{1}{2}b^2 & \text{in cases } I_2, II_2 \\ \text{(iii)} & c = -\tfrac{1}{2}b^2 & \text{in cases } I_3, II_3. \end{array}$$

It follows from the equivalence theorem of **2** that all surfaces which come under cases I_1 and II_1, or cases I_2 and II_2 for the same b, or cases I_3 and II_3 for the same b are conformally equivalent in R_3. It is only necessary to identify one of the surfaces in each category in order to obtain a complete classification of all isothermic gradient surfaces. This is done below:

(i) $c = 0$. For a surface belonging to II_1, (7.14) becomes, after using (7.37), (7.38), (7.19), (7.20) and (7.21),

$$ds^2 = du^2 + dv^2$$
$$b_{ij}dx^i dx^j = -2Q(v)dv^2.$$

It is clear from these forms that the surface V_2 is developable, that v denotes ordinary distance along one family of lines of curvature and that $-2Q(v)$ is the principal normal curvature of V_2 along each line of the family. Hence V_2 is a right cylindrical surface whose directrix is the plane curve defined intrinsically by

$$(7.51)\qquad k = -2(s).$$

Here k and s are the curvature and arc length of the directrix.

(ii) $c = \frac{1}{2}b^2$. A surface of type II_2 has the fundamental forms

$$(7.52)\qquad \begin{aligned} ds^2 &= (1/h^2)(du^2 + dv^2) \\ b_{ij}dx^i dx^j &= (f/h^2)du^2 + (f/h^2 - 2Q/h)dv^2 \end{aligned}$$

where $f(v), h(v)$ obey (7.22) and (7.40). Since the coefficients in (7.52) are all functions of v only, comparison with a surface of revolution is suggested. Now the equations of any surface of revolution are

$$y^1 = \omega\cos\alpha, \qquad y^2 = \omega\sin\alpha, \qquad y^3 = \nu(\omega).$$

From these equations, we obtain the following fundamental forms

$$(7.53)\qquad \begin{aligned} ds^2 &= \omega^2 d\alpha^2 + \sec^2\theta\cdot d\omega^2 \\ b_{ij}dx^i dx^j &= \omega\sin\theta\cdot d\alpha^2 + \sec\theta\cdot(d\theta/d\omega)\cdot d\omega^2 \end{aligned}$$

where $\theta(\omega) = \arctan(d\nu/d\omega)$. We introduce the isothermal parameter χ by the equation

$$\chi = \int (\sec\theta/\omega)\, d\omega. \tag{7.54}$$

If (7.52) and (7.53) refer to the same surface

$$u = d\alpha, \qquad v = d\chi \tag{7.55}$$

where d is a constant. Equating the coefficients of the respective fundamental forms of (7.52) and (7.53), we obtain the equations

$$h = d/\omega, \qquad f = \sin\theta/\omega \tag{7.56}$$

$$\omega^2 \cos\theta(d\theta/d\omega) = d^2(f/h^2 - 2Q/h). \tag{7.57}$$

From (7.54), (7.55) and (7.56), we obtain

$$dh/dv = \frac{1}{c}\frac{dh}{d\chi} = (1/c)\omega\cos\theta(dh/d\omega) = -\cos\theta/\omega$$

$$\begin{aligned} df/dv &= (1/c)\omega\cos\theta(df/d\omega) \\ &= (1/c)[\cos^2\theta(d\theta/d\omega) - (1/\omega)\cos\theta\sin\theta]. \end{aligned} \tag{7.58}$$

Now h and f satisfy (7.22) and (7.40). Substitution of (7.56) and (7.58) in (7.40) yields $d = 1/b$. Similarly (7.22) reduces to (7.57) which may be simplified by means of (7.56) to give

$$2dQ(d\chi) = \sin\theta - \omega\cos\theta(d\theta/d\omega). \tag{7.59}$$

If $Q(v)$ is given, equations (7.54) and (7.59) may always be solved for $\theta(\omega)$. Conversely, if $\theta(\omega)$ is given, (7.54) and (7.59) define $Q(v)$. Hence the V_2 defined by (7.52) is a surface of revolution and every surface of revolution is of type II_2.

(iii) $c = -\frac{1}{2}b^2$. A surface of type II_3 has the fundamental forms

$$\begin{aligned} ds^2 &= e^{2bu}(du^2 + dv^2) \\ b_{ij}dx^i dx^j &= -2Qe^{bu}dv^2. \end{aligned} \tag{7.60}$$

By means of the parameter transformation

$$u' = (1/b)e^{bu}, \qquad v' = bv \tag{7.61}$$

(7.60) becomes

$$\begin{aligned} ds^2 &= du'^2 + u'^2 dv'^2 \\ b_{ij}dx^i dx^j &= -(2/b)u'Q(v'/b)\cdot dv'^2. \end{aligned} \tag{7.62}$$

The first form of (7.62) is identical with that of the plane if the parameters are polar coördinates. Hence the surface V_2 defined by (7.62) is a developable surface whose lines of curvature are applicable to a plane system of polar

coördinates. Here u', v' correspond to the length of the radius vector and the vectorial angle respectively. This means that the V_2 defined by (7.62) must be a conical surface. The converse is also true. The line of curvature, $u' = 1/b$, may be considered as the intersection of the conical surface V_2 and the sphere whose center is at the point $u' = 0$ and whose radius is $1/b$. Classical results of differential geometry lead to the conclusion that the geodesic curvature of the intersection curve relative to the sphere is the normal curvature in its direction relative to the conical V_2. But, using (7.62), the normal curvature of $u' = 1/b$ is given by

$$_{(2)}b = -2Q(v)/bu' = -2Q(v).$$

Also, from (7.61), since dv' is a small central angle, dv is an element of distance on the sphere. Hence V_2 is the conical surface whose vertex is at the center of the sphere of radius $1/b$ and whose directrix on this sphere is defined by the intrinsic equation

$$k = -2Q(s).$$

Here k and s are the geodesic curvature and arc length of the spherical curve. We summarize these results in the following theorem:

THEOREM 7.2. *Every isothermic gradient surface is conformally equivalent to a conical surface, cylindrical surface or a surface of revolution, and conversely.*

The generators of a cylindrical surface may be considered as circles which are all tangent to each other at the point at infinity. Similarly, the generators of a conical surface may be considered as circles which all have two common points—the vertex and the point at infinity. Consequently an isothermic gradient surface is of type (i), (ii), (iii) of (7.50) according as the members of the family of circular lines of curvature have 1, 0 or 2 common points. In particular, this topological classification applies to Dupin cyclides.

8. Gradient surfaces of constant conformal Gaussian curvature. We consider a gradient surface V_2 for which J is constant. Let the lines of curvature be parametric. Since V_2 is gradient, in accordance with the results of 4, we may choose the components of its conformal measure tensor as

$$X = e^{2Q}, \qquad Y = e^{-2Q}. \tag{8.1}$$

The condition (3.12) is satisfied in virtue of (8.1). Using (4.1) and (4.8), the remaining conditions (3.11), (3.13) and (3.14) may be written as

$$_{(1)}E + {}_{(2)}E = J + 1 \tag{8.2}$$

(8.3) $$_{(1)}E_v = -Q_v(E'+2)$$

(8.4) $$_{(2)}E_u = -Q_u(E'-2).$$

From (4.8) and (8.2),

(8.5) $$_{(1)}E = \tfrac{1}{2}[1+J+E'], \qquad _{(2)}E = \tfrac{1}{2}[1+J-E'].$$

Since J is constant, if these values are substituted in (8.3) and (8.4) we obtain

(8.6) $$E'_v = -2Q_v(E'+2)$$

(8.7) $$E'_u = \quad 2Q_u(E'-2).$$

We consider the possible solutions of (8.6) and (8.7).

(I) Suppose that Q is a constant. Then $J=0$ and V_2 is a Dupin cyclide.

(II) Suppose that Q is a non-constant function[4] $Q(v)$ of v only. Then, from (8.7), $E' = E'(v)$. Integration of (8.6) yields

(8.8) $$E' = c_1 e^{-2Q} - 2$$

where c_1 is a constant. The remaining condition, $J =$ constant, leads to

(8.9) $$e^{2Q} = -Jv^2 + c_2 v + c_3$$

where the c's are constants, after use of (3.7) and (3.10). We consider three cases:

(II_1) $J=0$. We may suppose that $c_2 \neq 0$, the other alternative having been considered in (I). Then the substitution of (8.9) in (8.1) followed by the parameter transformation

$$u' = c_2 u/2, \qquad v' = (2/c_2)(c_2 v + c_3)^{1/2}$$

permits us to write the first conformal fundamental form of V_2 as

(8.10) $$dS^2 = v^2 du^2 + dv^2.$$

The primes have been omitted in (8.10). The deviation vector and the remaining conformal fundamental forms are readily found by reference to (3.2), (3.5), (8.2) and (8.8). They are

(8.11) $$F_1 = 0 \qquad F_2 = -2/v$$
$$B_{ij}dx^i dx^j = v^2 du^2 - dv^2$$
$$E_{ij}dx^i dx^j = (-\tfrac{1}{2}v^2 + d)du^2 + (3/2 - d/v^2)dv^2$$

[4] The case $Q = Q(u)$ is equivalent to the one discussed here after a parameter transformation which interchanges the roles of u and v.

where d is an arbitrary constant.

(II_2) $J = a^2$. In this case, we may write (8.9) as

$$e^{2Q} = c_4^2 - (av + c_5)^2. \tag{8.12}$$

If we substitute (8.12) in (8.1) and employ the parameter substitution

$$u' = c_4 u, \qquad v' = (1/a) \arcsin [(av + c_5)/c_4]$$

the first conformal fundamental form of V_2 becomes [5]

$$dS^2 = \cos^2 av du^2 + dv^2 \tag{8.13}$$

after dropping the primes. The other forms and F_i are found as in (II_1) and are

$$\begin{aligned} B_{ij}dx^i dx^j &= \cos^2 av du^2 - dv^2 \\ E_{ij}dx^i dx^j &= [\tfrac{1}{2}(a^2 - 1)\cos^2 av + d]du^2 \\ &\quad + [\tfrac{1}{2}(a^2 + 3) - d\sec^2 av]dv^2 \\ F_1 &= 0, \qquad F_2 = 2a \tan av. \end{aligned} \tag{8.14}$$

(II_3) $J = -a^2$. We may now write (8.9) as

$$e^{2Q} = (av + c_5)^2 \pm c_4^2. \tag{8.15}$$

We substitute (8.15) in (8.1) and then use the parameter transformation

(i) $u' = c_4 u, \qquad v' = (1/a) \operatorname{arc\,cosh} [(av + c_5)/c_4]$

(ii) $u' = c_4 u, \qquad v' = (1/a) \operatorname{arc\,sinh} [(av + c_5)/c_4]$

(iii) $u' = u, \qquad v' = (1/a) \log (av + c_5)$

according as (8.15) contains the plus sign or minus sign or $c_4 = 0$. In this manner, we obtain three possible forms:

$$\begin{aligned} &\text{(i)} && dS^2 = \cosh^2 av du^2 + dv^2 \\ &\text{(ii)} && dS^2 = \sinh^2 av du^2 + dv^2 \\ &\text{(iii)} && dS^2 = e^{2av} du^2 + dv^2 \end{aligned} \tag{8.16}$$

after dropping the primes. The other forms and F_i are

$$\begin{aligned} B_{ij}dx^i dx^j &= X du^2 - dv^2 \\ E_{ij}dx^i dx^j &= [-\tfrac{1}{2}(1 + a^2)X + d]du^2 + [\tfrac{1}{2}(3 - a^2) - dX^{-1}]dv^2 \\ F_1 &= 0, \qquad F_2 = -(\log X)v \end{aligned} \tag{8.17}$$

where X equals $\cosh^2 av$, $\sinh^2 av$, e^{2av} in cases (i), (ii), (iii), respectively.

---

[5] Equation (8.13) may also be written

$$dS^2 = \sin^2 av \cdot du^2 + dv^2.$$

A comparison of (7.10) with (8.11), (8.14) and (8.17) respectively reveals that

$$(8.18)\qquad d = \begin{cases} \frac{1}{2}(2c-1), & J=0 \\ \frac{1}{2}(2c-a^2), & J=a^2 \\ \frac{1}{2}(2c+a^2), & J=-a^2, \text{ case (i)} \\ \frac{1}{2}(2c-a^2), & J=-a^2, \text{ case (ii)} \\ c, & J=-a^2, \text{ case (iii)}. \end{cases}$$

(III) Next let us suppose that both variables u, v are effectively present in Q so that $Q_u \cdot Q_v \neq 0$. It follows from (8.6) and (8.7) that E' cannot be 2 or -2 and indeed must be a function of both u and v. Integration of (8.6) and (8.7) leads to

$$E'+2 = g_1(u)\cdot e^{-2Q}, \qquad E'-2 = h_1(v)\cdot e^{2Q}.$$

Since we obtain

$$E' = [g_1h_1+4]^{\frac{1}{2}}$$

from these equations it follows that neither g_1 nor h_1 is constant. We also find that

$$e^{2Q} = (E'-2)/h_1 = g_1/(E'+2).$$

Hence the first conformal fundamental form of V_2 may be written as

$$(8.19)\qquad dS^2 = \frac{[g_1h_1+4]^{\frac{1}{2}}-2}{h_1}\,du^2 + \frac{[g_1h_1+4]^{\frac{1}{2}}-2}{g_1}\,dv^2.$$

After the parameter transformation

$$u' = \tfrac{1}{2}g_1(u), \qquad v' = \tfrac{1}{2}h_1(v)$$

(8.19) becomes

$$(8.20)\qquad dS^2 = \frac{[uv+1]^{\frac{1}{2}}-1}{vg_2(u)}\,du^2 + \frac{[uv+1]^{\frac{1}{2}}+1}{uh_2(v)}\,dv^2$$

where we have written

$$g_2(u') = \tfrac{1}{4}g'_1{}^2[u(u')], \qquad h_2(v') = \tfrac{1}{4}h'_1{}^2[v(v')]$$

and dropped the primes.

It remains to impose the condition that the V_2 having (8.20) as its first conformal fundamental form has a constant value for J (or K). We shall prove that it is impossible to satisfy this condition. From (3.7) and (8.20),

$$(8.21)\qquad \begin{aligned} K = {} & \frac{(3u^2v^2+12uv+8)}{4(uv+1)^{3/2}}\cdot(g_2/u^3+h_2/v^3) - \frac{(uv+2)}{4(uv+1)^{\frac{1}{2}}}\cdot(g'_2/u^2+h'_2/v^2) \\ & + 2g_2/u^3 - 2h_2/v^3 - g'_2/2u^2 + h'_2/2v^2. \end{aligned}$$

We define $g(u)$, $h(v)$ by

$$g = g_2/u^3, \qquad h = h_2/v^3.$$

Then

$$g'_2/u^2 = 3g + ug', \qquad h'_2/v^2 = 3h + vh'.$$

Hence (8.21) may be written as

$$K = \frac{3uv+2}{4(uv+1)^{3/2}}(g+h) - \frac{uv+2}{4(uv+1)^{1/2}}(ug'+vh') \tag{8.22}$$
$$-g+h-ug'+vh'.$$

The first two expressions in the right member of this equation are of the form

$$e(u,v) = f(z)[u^n g^{(n)} + v^n h^{(n)}]$$

where (n) denotes n-fold differentiation and $z = uv$. Successive differentiation of this equation gives

$$e_{uv} = (zf'' + (1+n)f')[u^n g^{(n)} + v^n h^{(n)}] \tag{8.23}$$
$$+ f'[u^{n+1}g^{(n+1)} + v^{n+1}h^{(n+1)}]$$

where the prime indicates differentiation with respect to z. We differentiate (8.22) successively with respect to u and v and apply (8.23). After multiplication by $16(z+1)^{7/2}$, we obtain

$$(3z^2-12z)(g+h) - (3z^3+15z^2+12z)[ug^{(1)}+vh^{(1)}]$$
$$-(2z^3+4z^2+2z)[u^2g^{(2)}+v^2h^{(2)}] = 0.$$

If we apply (8.23) to this equation three more times we are led to the linear differential equation

$$72[ug^{(1)}+vh^{(1)}] + 228[u^2g^{(2)}+v^2h^{(2)}] + 156[u^3g^{(3)}+v^3h^{(3)}]$$
$$+33[u^4g^{(4)}+v^4h^{(4)}] + 2[u^5g^{(5)}+v^5h^{(5)}] = 0.$$

From this equation, we conclude that

$$72ug^{(1)} + 228u^2g^{(2)} + 156u^3g^{(3)} + 33u^4g^{(4)} + 2u^5g^{(5)} = c_1$$
$$72vh^{(1)} + 228v^2h^{(2)} + 156v^3h^{(3)} + 33v^4h^{(4)} + 2v^5h^{(5)} = -c_1.$$

Integration of these differential equations by the customary methods leads to

$$g(u) = (c_1/6)\log u + c_2 + c_3u^{-1/2} + c_4u^{-1} + c_5u^{-2} + c_6u^{-3}$$

and a similar solution for $h(v)$. But substitution of these solutions in (8.22) readily leads to a contradiction. Hence functions $g(u), h(v)$ which satisfy

(8.22) do not exist. This means that there are no surfaces which obey the requirements of case (III). Since the first conformal fundamental form for each surface which occurs under case (I) or case (II) satisfies (7.9), it follows from Theorem 7.1 that these surfaces are all isothermic gradient. We state these results in the following theorem:

THEOREM 8.1. *If a gradient surface has constant conformal Gaussian curvature J, then it is isothermic. If $J = 0$, the surface is a Dupin cyclide or its conformal fundamental forms are defined by* (8.10) *and* (8.11) *in some parameter system. If J equals $+a^2$ or $-a^2$ its conformal fundamental forms are defined by* (8.13) *and* (8.14) *or by* (8.16) *and* (8.17), *respectively, in some parameter system.*

As noted in **6**, it follows from (3.11) and (4.1) that a surface V_2 whose principal deviations are each constant has constant conformal Gaussian curvature. Consequently if V_2 is also a gradient surface, its conformal fundamental forms are defined by Theorem 8.1. In this case, unless V_2 is a Dupin cyclide, the only constant values of ${}_{(1)}E$ and ${}_{(2)}E$ are obtained by equating d to zero in (8.11), (8.14) and (8.17). Hence *a gradient surface having constant principal deviations is either a Dupin cyclide or its principal deviations are*

$$(8.24) \qquad {}_{(1)}E = -1/2 + J/2, \qquad {}_{(2)}E = 3/2 + J/2.$$

According to the results of **7**, each of these surfaces is conformally equivalent to a cylindrical surface, surface of revolution or conical surface according as c equals, is greater than or is less than zero. Since $d = 0$, it follows from (8.18) that a surface (which is not a Dupin cyclide) having constant principal deviations is conformally equivalent to a surface of revolution if J is zero or positive. If J is negative, the surface is conformal to a conical surface, surface of revolution or cylindrical surface in cases (i), (ii) or (iii) of (8.16) respectively.

As an example of such a surface, we mention the catenoid referred to in **6** for which $J = 1$. We also note the right cylindrical surface whose directrix is the equiangular spiral $\rho = e^{(a/2)\theta}$. The intrinsic equation of the spiral is $k = 2/as$. Comparison with (7.51) shows that the first conformal fundamental form of the cylindrical surface is

$$dS^2 = (1/a^2v^2)(du^2 + dv^2)$$

which is equivalent to (iii) of (8.16) after the introduction of isothermal parameters in the latter.

9. Isothermic surfaces of constant conformal Gaussian curvature. If

V_2 is isothermic, the first and second conformal fundamental forms may be written as

$$\begin{aligned} dS^2 &= e^{2Q}(du^2 + dv^2) \\ B_{ij}dx^i dx^j &= e^{2Q}(du^2 - dv^2). \end{aligned} \tag{9.1}$$

In this parameter system, from (3.7) and (3.10),

$$J = -e^{-2Q}(Q_{uu} + Q_{vv}). \tag{9.2}$$

The conditions (3.11) to (3.14) become

$$E_{11} + E_{22} = e^{2Q}(J+1) \tag{9.3}$$

$$E_{12} = -2Q_{uv} \tag{9.4}$$

$$\partial E_{11}/\partial v = -2Q_{uuv} + e^{2Q}Q_v(J-1) \tag{9.5}$$

$$\partial E_{22}/\partial u = -2Q_{uvv} + e^{2Q}Q_u(J-1) \tag{9.6}$$

respectively. We also assume that J is constant. Then, as a consequence of (9.3) and (9.6), we obtain

$$\partial E_{11}/\partial u = 2Q_{uvv} + e^{2Q}Q_u(J+3). \tag{9.7}$$

The integrability condition of (9.5) and (9.7) may be written as

$$[Q_{uu} + Q_{vv}]_{uv} + 2e^{2Q}(2Q_uQ_v + Q_{uv}) = 0$$

or, using (9.2),

$$e^{2Q}(1-J)(2Q_uQ_v + Q_{uv}) = 0. \tag{9.8}$$

We first suppose that $J \neq 1$. Then (9.8) yields

$$e^{2Q} = Q_1(u) + Q_2(v). \tag{9.9}$$

By virtue of (9.9), (9.5) and (9.6) may be integrated. This gives

$$\begin{aligned} E_{11} &= -2Q_{uu} + \tfrac{1}{2}(J-1)e^{2Q} + g(u), \\ E_{22} &= -2Q_{vv} + \tfrac{1}{2}(J-1)e^{2Q} + h(v). \end{aligned} \tag{9.10}$$

Substitution of these results in (9.2) and use of (9.9) leads to

$$g(u) = c - 2(J-1)Q_1, \qquad h(v) = -c - 2(J-1)Q_2.$$

If these values are substituted in (9.10), we obtain

$$\begin{aligned} E_{11} &= -2Q_{uu} + \tfrac{1}{2}(J-1)(Q_2 - 3Q_1) + c \\ E_{22} &= -2Q_{vv} + \tfrac{1}{2}(J-1)(Q_1 - 3Q_2) - c. \end{aligned} \tag{9.11}$$

It only remains to satisfy (9.2). In this way, from (9.2) and (9.9),

$$Q_1''Q_1 - Q_1'^2 + Q_2''Q_2 - Q_2'^2 + Q_2Q_1'' + Q_1Q_2'' = -2J(Q_1+Q_2)^2 \tag{9.12}$$

where the primes indicate differentiation, with respect to the independent variable, of each function. If $Q'_1Q'_2 = 0$, we may suppose, after a parameter change, that $Q_1 = 0$. Then (9.12) becomes

$$Q_2''Q_2 - Q'^2_2 + 2JQ_2^3 = 0. \tag{9.13}$$

In accordance with Theorem 7.1, the surfaces for which $Q_1 = 0$ must be isothermic gradient surfaces with constant J. They are discussed in **8**. The first conformal fundamental forms of these surfaces, when the parameters are isothermal, may be derived from (9.13) or, more simply, by restating (6.1), (8.10), (8.13) and (8.16) in an isothermal parameter system. If $Q'_1Q'_2 \neq 0$, successive differentiation of (9.12) with respect to u and v and division by $Q'_1Q'_2$ gives

$$Q_1'''/Q'_1 + 12JQ_1 + Q_2'''/Q_2' + 12JQ_2 = 0$$

which, after separation and integration, leads to

$$\begin{aligned} Q_1'' &= -6JQ_1^2 + cQ_1 + c_1 \\ Q''_2 &= -6JQ_2^2 - cQ_2 + c_2 \end{aligned} \tag{9.14}$$

where the c's are constants.

We consider various cases:

(I) $J = 0$. The solutions of (9.13) define first conformal fundamental forms which are equivalent to (6.1) and (8.10). These forms are

$$\begin{aligned} dS^2 &= du^2 + dv^2 \\ dS^2 &= e^{2v}(du^2 + dv^2) \end{aligned} \tag{9.15}$$

respectively.

We now consider the solution of (9.14) and assume that $c = 0$. Then

$$Q_1 = (c_1/2)u^2 + c_3u + c_4, \qquad Q_2 = (c_2/2)v^2 + c_5v + c_6.$$

Substitution of these results in (9.12) gives the additional conditions

$$c_1 = c_2, \qquad 2c_1(c_4 + c_6) = c_3^2 + c_5^2.$$

Hence

$$e^{2Q} = (c_1/2)[(u + c_7)^2 + (v + c_8)^2].$$

After the parameter transformation

$$u' = (c_1/2)^{1/4}(u + c_7), \qquad v' = (c_1/2)^{1/4}(v + c_8)$$

and dropping primes, (9.1) becomes

$$dS^2 = (u^2 + v^2)(du^2 + dv^2). \tag{9.16}$$

Now suppose that $c \neq 0$ in (9.14). If $c = b^2$, integration of (9.14) gives

$$Q_1 = \begin{cases} c_3 \cosh(bu + c_4) - c_1/b^2 \\ c_3 \sinh(bu + c_4) - c_1/b^2 \\ c_3 e^{bu} - c_1/b^2 \end{cases}$$

$$Q_2 = \begin{cases} c_5 \cos(bv + c_6) + c_2/b^2 \\ c_5 \sin(bv + c_6) + c_2/b^2. \end{cases}$$

The solutions for Q_2 are equivalent to each other while those for Q_1 are not. If the first (or second) solution of each set is substituted in (9.12), we obtain

$$c_1 = c_2, \qquad c_3 = \pm c_5.$$

After the parameter transformation

$$u' = bu + c_4, \qquad v' = bv + c_6$$

and writing $c_3/b^2 = a^2$ and dropping primes, the forms (9.1) corresponding to these solutions are

$$(9.17) \qquad \begin{aligned} dS^2 &= a^2(\cosh u \pm \cos v)(du^2 + dv^2) \\ dS^2 &= a^2(\sinh u \pm \sin v)(du^2 + dv^2). \end{aligned}$$

If the third solution for Q_1 and either solution for Q_2 is substituted in (9.12), we find that $c_5 = 0$. Hence this solution is excluded since Q_2 must actually involve v. Finally, we note that if $c = -b^2$ in (9.14), u and v interchange rôles so that the resulting solution is equivalent to (9.17) after a change of parameter.

(II) $J = a^2$. The first conformal fundamental form defined by the solutions of (9.13) is equivalent to (8.13). In this way, we obtain

$$(9.18) \qquad dS^2 = \operatorname{sech}^2 av(du^2 + dv^2).$$

We now consider the solution of (9.14). Integration leads to

$$(9.19) \qquad \begin{aligned} Q'_1{}^2 &= -4a^2Q_1{}^3 + cQ_1{}^2 + 2c_1Q_1 + c_3 \\ Q'_2{}^2 &= -4a^2Q_2{}^3 - cQ_2{}^2 + 2c_2Q_2 + c_4 \end{aligned}$$

where the c's are constants. If these equations and (9.14) are substituted in (9.12), we find that

$$c_1 = c_2, \qquad c_3 = -c_4$$

so that, after the substitution

$$Q_1 = Q_3 + c/12a^2, \qquad Q_2 = Q_4 - c/12a^2$$

(9.19) may be written as

$$Q'_3{}^2 = -a^2[4Q_3{}^2 - g_2Q_3 - g_3]$$
$$Q'_4{}^2 = -a^2[4Q_4{}^2 - g_2Q_4 + g_3]$$

where the g's are constants. The solutions of these equations, after translation of u and v, if necessary, are

$$(9.20) \qquad \begin{aligned} Q_1 &= \mathfrak{P}(iau \mid g_2, g_3) + c/12a^2 \\ Q_2 &= \mathfrak{P}(iav \mid g_2, -g_3) - c/12a^2 \end{aligned}$$

where $\mathfrak{P}(x \mid g_2, g_3)$ is the Weierstrass $\mathfrak{P}$-function of x with the invariants g_2, g_3. Since $\mathfrak{P}(x)$ is an even function and g_2, g_3 are real, the functions Q_1 and Q_2 defined by (9.20) are real for real values of u and v. The first conformal fundamental form corresponding to (9.20) is

$$(9.21) \qquad dS^2 = [\mathfrak{P}(iau \mid g_2, g_3) + \mathfrak{P}(iav \mid g_2, g_3)](du^2 + dv^2).$$

(III) $J = -a^2$. The solutions of (9.13) define first conformal fundamental forms which are equivalent to (8.16). These forms are

$$(9.22) \qquad \begin{aligned} &\text{(i)} && dS^2 = \sec^2 av(du^2 + dv^2) \\ &\text{(ii)} && dS^2 = \operatorname{csch}^2 av(du^2 + dv^2) \\ &\text{(iii)} && dS^2 = (1/a^2v^2)(du^2 + dv^2). \end{aligned}$$

If we proceed wth (9.14) as in (II) above, we obtain

$$(9.23) \qquad dS^2 = [\mathfrak{P}(au \mid g_2, g_3) + \mathfrak{P}(av \mid g_2, -g_3)](du^2 + dv^2)$$

as the first conformal fundamental form corresponding to the solution.

(IV) We now assume that $J = 1$. Then (9.8) is true. In this case, any solution $Q(u, v)$ of (9.2) with $J = 1$ defines the first conformal fundamental form of an actual surface. The components of E_{ij} are then defined by (9.4) and by

$$E_{11} = -2Q_{uu} + c, \qquad E_{22} = -2Q_{vv} - c.$$

These last equations are special cases of (9.11).

We summarize these results as follows:

THEOREM 9.1. *An isothermic surface of constant conformal curvature J ($J \neq 1$) is either a gradient surface*[6] *or its first conformal fundamental form is defined by* (9.16) *or* (9.17) *if $J = 0$, by* (9.21) *if $J = a^2$ and by* (9.23) *if $J = -a^2$. If $J = 1$, any solution of* (9.2) *defines the first con-*

[6] In this case, its conformal fundamental forms are given by Theorem 8.1.

formal fundamental form of a surface. The components of the other conformal fundamental forms in all cases are defined by (9.1), (9.4) *and* (9.11).

10. Surfaces with indeterminate lines of deviation. If a surface V_2 is homogeneous with respect to the deviation tensor,

$$E_{ij} = (E/2)G_{ij}. \tag{10.1}$$

In this case, every direction is a deviation direction. We say V_2 has *indeterminate lines of deviation.* It follows from Theorem 4.2 that V_2 is a gradient surface. Hence the components of its conformal measure tensor may be chosen as (8.1) and the remaining conditions are (8.2), (8.3) and (8.4). However, in virtue of (4.1), (4.8) and (10.1)

$$_{(1)}E = {}_{(2)}E = E/2, \qquad E' = 0. \tag{10.2}$$

Therefore (8.3) and (8.4) become

$$E_v = -4Q_v, \qquad E_u = 4Q_u. \tag{10.3}$$

The integrability conditions of these equations are $Q_{uv} = 0$ so that Q may be written as

$$Q = \log Q_2(v) - \log Q_1(u); \qquad e^Q = Q_2/Q_1. \tag{10.4}$$

As a result of (10.3) and (10.4),

$$E = -4 \log Q_1Q_2 + c \tag{10.5}$$

where c is a constant. If (10.2) and (10.5) are substituted in (8.2), we obtain

$$J = -4 \log Q_1Q_2 + c - 1.$$

But, from (3.7) and (3.10),

$$J = -\tfrac{1}{2}[Q_1''/Q_2 + Q_2''/Q_1]$$

where the primes indicate differentiation of the functions with respect to their independent variables. From the last two equations,

$$Q_1''/Q_2 + Q_2''/Q_1 = 8 \log Q_1Q_2 - \tfrac{1}{2}(c-1). \tag{10.6}$$

Successive differentiation of (10.6) with respect to u and v yields

$$Q_1'''(1/Q_2)' + Q_2'''(1/Q_1)' = 0.$$

If $Q'_1Q'_2 \neq 0$, this equation leads to two separate equations which integrate as

$$\begin{aligned} Q_1'' &= c_1 \cdot 1/Q_1 + c_2 \\ Q_2'' &= -c_1(1/Q_2) + c_3 \end{aligned} \tag{10.7}$$

where the c's are constants. Substitution of (10.7) in (10.6) leads to

$$c_2/Q_2 + c_3/Q_1 = 8 \log Q_1 + 8 \log Q_2 - \tfrac{1}{2}(c-1)$$

which is impossible.

Therefore $Q'_1 Q'_2 = 0$ and we may suppose, after a parameter change, that $Q_1 = 1$. Hence V_2 is isothermic. The only condition on Q_2 is (10.6) which becomes

$$Q_2'' = 8 \log Q_2 - \tfrac{1}{2}(c-1). \tag{10.8}$$

Any solution of (10.8) together with (10.4) and (8.1) define the components of the first conformal fundamental form of the surface. The only solution of (10.8) for which E is constant is $Q_2 = 1$, $c = 1$ as follows from (10.5). In this case, $E = 1$. The corresponding V_2 is a Dupin cyclide. A comparison of (10.1) with (6.8) shows that V_2 is conformally equivalent to a torus for which $a = \sqrt{2}d$. This torus is characterized by the fact that the tangents drawn from the origin to its generating circle are orthogonal. We state some of these results in the following theorem:

THEOREM 10.1. *Every surface with indeterminate lines of deviation is isothermic. The deviation of the surface is constant if and only if it is conformally equivalent to a torus whose generating circle subtends a right angle at the projection of its center on the axis of rotation.*

11. Gradient hypersurfaces. We call a hypersurface V_n in a Euclidean space R_{n+1} a *gradient hypersurface* if F_i is a gradient on V_n. In this case, there exists a single parameter family of subspaces in V_n, whose equations may be written as $F = \text{constant}$, to which the vector F_i is normal. In this section we derive an analogue of Theorem 4.2 for V_n. The necessary and sufficient condition that F_i be a gradient $F_{;i}$ is that

$$F_{j:k} - F_{k:j} = 0. \tag{11.1}$$

According to (2.22), (11.1) is equivalent to

$$B_{kj}E_k{}^h - B_{hk}E_j{}^h = 0. \tag{11.2}$$

From (2.9) and (2.15), it follows that the n mutually orthogonal principal directions ${}_{(1)}L_i, {}_{(2)}L_i, \cdots, {}_{(n)}L_i$ determined by the tensor B_{ij} relative to G_{ij}

coincide with those determined by b_{ij} relative to g_{ij}. Therefore they are tangent to the lines of curvature of V_n. We also refer to them as the *curvature directions*. The corresponding roots ${}_{(1)}B, {}_{(2)}B, \cdots, {}_{(n)}B$ of the determinantal equation $|B_{ij} - \rho G_{ij}| = 0$ are the *principal conformal normal curvatures* of V_n. The tensor B_{ij} then has the canonical representation

$$B_{ij} = {}_{(1)}B\,{}_{(1)}L_i\,{}_{(1)}L_j + {}_{(2)}B\,{}_{(2)}L_i\,{}_{(2)}L_j + \cdots + {}_{(n)}B\,{}_{(n)}L_i\,{}_{(n)}L_j. \tag{11.3}$$

Now the tensor equations (11.2) are equivalent to the set of scalar equations obtained by multiplying (11.2) by ${}_{(g)}L^i\,{}_{(l)}L^k$; $g, l = 1, 2, \cdots, n$. Thus (11.2) becomes

$$[{}_{(g)}B - {}_{(l)}B]E_{hj}\,{}_{(g)}L^h\,{}_{(l)}L^j = 0 \tag{11.4}$$

after use of (11.3). There are two ways in which (11.4) may be satisfied. (1) Suppose that ${}_{(g)}B$ is a simple principal conformal normal curvature; that is, ${}_{(g)}B \neq {}_{(l)}B$ for a fixed g and any l not equal to g. Then (11.4) implies that ${}_{(g)}L^h$ is a principal direction of E_{hj} relative to G_{hj}. This means that ${}_{(g)}L^h$ is a deviation direction. (2) Suppose that ${}_{(g)}B$ is a multiple principal conformal normal curvature so that ${}_{(g)}B = {}_{(l)}B$ for $g, l = 1, 2, \cdots, w$. Then the known theory and (11.4) imply that w principal directions of E_{ij} lie in the w-dimensional linear vector space determined by ${}_{(1)}L^i, {}_{(2)}L^i, \cdots, {}_{(w)}L^i$. Due to the symmetry of (11.2), the rôles of E_{ij}, B_{ij} may be interchanged in the above remarks. This completes the proof of the following analogue of Theorem 4.2:

THEOREM 11.1. *If V_n is a gradient hypersurface then* (1) *the curvature direction corresponding to a simple principal conformal normal curvature is also a deviation direction* (2) *the w-dimensional linear vector space of curvature directions corresponding to a principal conformal normal curvature of multiplicity w contains w deviation directions, and conversely.*

The above theorem remains valid if the words "principal conformal normal curvature," "curvature direction" are interchanged with "principal deviation," "deviation direction" respectively.

From (2.19) and (2.20), we obtain

$$K_{ij} = B_{ik}B_{hj}G^{hk} + (2-n)E_{ij} - EG_{ij} \tag{11.5}$$

after multiplication by G^{hk} and summation on repeated indices. The principal

directions determined by the *conformal Ricci (surface) tensor* K_{ij} are the *conformal Ricci (surface) directions.* The tensor $B_{ik}B_{hj}G^{hk}$ has the canonical representation

$$B_{ik}B_{hj}G^{hk} = {}_{(1)}B^2\,{}_{(1)}L_i\,{}_{(1)}L_j + {}_{(2)}B^2\,{}_{(2)}L_i\,{}_{(2)}L_j + \cdots + {}_{(n)}B^2\,{}_{(n)}L_i\,{}_{(n)}L_j$$

as follows from (11.3). Hence the principal directions determined by $B_{ik}B_{hj}G^{hk}$ are tangent to the lines of curvature of V_n. As a consequence of this remark and (11.5), we have

THEOREM 11.2. *If a vector is a member of two of the following sets*

(1) *the curvature directions of* V_n

(2) *the deviation directions of* V_n

(3) *the conformal Ricci (surface) directions of* V_n

and n exceeds 2, it is also a member of the third.

In particular, this theorem has obvious applications to gradient hypersurfaces, as may be seen by reference to Theorem 11.1.

12. Hypersurfaces with indeterminate lines of deviation. A V_n has *indeterminate lines of deviation* if and only if

$$E_{ij} = (E/n)G_{ij}. \tag{12.1}$$

If (12.1) is substituted in (2.22), we obtain (11.1). Hence *any* V_n *with indeterminate lines of deviation is a gradient hypersurface.* We may therefore write F_i as $F_{:i}$. Substitution of (12.1) in (2.23) yields

$$E_{:k}G_{ij} - E_{:j}G_{ik} = n[B_{ij}F_{:k} - B_{ik}F_{:j}]. \tag{12.2}$$

Multiplication of (12.2) by G^{ij} and summation over i, j leads to

$$(n-1)E_{:k} = -nB_{ik}F_{:j}G^{ij} \tag{12.3}$$

after account is taken of (2.19). It follows from (12.3) and (11.3) that V_n *has constant deviation if and only if either* (1) $F_{:i} = 0$ or (2) $F_{:i} \neq 0$ *and any subspace,* $F =$ *constant, contains each line of curvature with which it has*

a point in common and whose corresponding principal conformal normal curvature differs from zero. The second alternative implies that the rank of $\| B_{ij} \| < n$ so that at least one of the principal conformal normal curvatures vanishes throughout V_n. This case cannot be realized if $n = 2$ as follows from (3.2).

Conformal differentiation of (12.3) with respect to x^k gives

$$(12.4) \qquad (n-1)E_{:ik} = -n[B_{ik:\lambda} F^\lambda + B_{ik} F_{:j} G^{ij}].$$

Since $E_{:ik} - E_{:ki} = 0$, (12.4) in conjunction with (2.21) leads to

$$(12.5) \qquad [B_{ik} F_{:j} - B_{ij} F_{:k}] G^{ij} = 0.$$

Equations (12.5) are exactly analogous to (11.2) so that a theorem like Theorem 11.1 exists in this case with $F_{:ij}$ occupying the rôle that E_{ij} plays in that theorem.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

BIBLIOGRAPHY.

1. L. P. Eisenhart, *Riemannian Geometry*, 1926, Princeton.
2. A. Fialkow, "Conformal theory of curves," *Proceedings of the National Academy of Sciences*, vol. 26, (1940), pp. 437-439.
3. ———, "Conformal theory of curves," *Transactions of the American Mathematical Society*, vol. 51 (1942), pp. 435-501.
4. ———, "Conformal differential geometry of a subspace," *Transactions of the American Mathematical Society*, vol. 56 (1944), pp. 309-433.

American Journal of Mathematics. Edited by R. BRAUER, F. D. MURNAGHAN, H. WEYL, H. WHITNEY and A. WINTNER. Quarterly. 8vo. Volume LXVII in progress. $7.50 per volume. (Foreign postage, fifty cents.)

American Journal of Philology. Edited by B. D. MERITT, H. CHERNISS, R. H. HAYWOOD, K. MALONE, H. T. ROWELL, and D. M. ROBINSON, honorary editor. Quarterly. 8vo. Volume XLVI in progress. $5 per volume. (Foreign postage twenty-five cents.)

Bulletin of the History of Medicine. Edited by HENRY E. SIGERIST. Monthly. Volume XVIII in progress. 8vo. Subscription $5 per year. (Foreign postage, fifty cents.)

Bulletin of the Johns Hopkins Hospital. Edited by L. EMMETT HOLT, JR. Monthly. Volume LXXVII in progress. 8vo. Subscription $6 per year. (Foreign postage, fifty cents.)

ELH. A Journal of English Literary History. Edited by E. T. NORRIS (managing editor) and D. C. ALLEN, G. E. BENTLEY, R. D. HAVENS, L. HOWARD, C. P. LYONS, and E. WASSERMAN. Quarterly. 8vo. Volume XII in progress. $2.50 per volume.

Hesperia. Edited by WILLIAM KURRELMEYER and KEMP MALONE. 8vo. Thirty-two numbers have appeared.

Human Biology: a record of research. LOWELL J. REED, Editor. Quarterly. 8vo. Volume XVII in progress. $5 per volume. (Foreign postage, thirty-five cents.)

Johns Hopkins Studies in Romance Literatures and Languages. H. C. LANCASTER, Editor. 8vo. Sixty-four numbers have been published.

Johns Hopkins University Circular, including the President's Report and Catalogue of the School of Medicine. Eight times yearly. 8vo. $1 per year.

Johns Hopkins University Studies in Archaeology. DAVID M. ROBINSON, Editor. 8vo. Thirty-five volumes have appeared.

Johns Hopkins University Studies in Education. FLORENCE E. BAMBERGER, Editor. 8vo. Thirty-four numbers have appeared.

Johns Hopkins University Studies in Geology. JOSEPH T. SINGEWALD, JR., Editor. 8vo. Fourteen numbers have been published.

Johns Hopkins University Studies in Historical and Political Science. Under the direction of the Departments of History, Political Economy and Political Science. 8vo. Volume LXII in progress. $5 per volume.

Modern Language Notes. Edited by H. C. LANCASTER, W. KURRELMEYER, R. D. HAVENS, K. MALONE, H. SPENCER, C. S. SINGLETON, C. R. ANDERSON and D. C. ALLEN. Eight times yearly. 8vo. Volume LX in progress. $5 per volume. (Foreign postage, fifty cents.)

Reprint of Economic Tracts. Founded by J. H. HOLLANDER. Five series are complete.

Terrestrial Magnetism and Atmospheric Electricity. Founded by LOUIS A. BAUER; conducted by J. A. FLEMING with the coöperation of eminent investigators. Quarterly. 8vo. Volume L in progress. $3.50 per volume.

Walter Hines Page School of International Relations. Nine volumes have been published.

A complete list of publications will be sent upon request

THE THEORY OF GROUP REPRESENTATIONS

By Francis D. Murnaghan

We have attempted to give a quite elementary and self-contained account of the theory of group representations with special reference to those groups (particularly the symmetric group and the rotation group) which have turned out to be of fundamental significance for quantum mechanics (especially nuclear physics). We have devoted particular attention to the theory of group integration (as developed by Schur and Weyl); to the theory of two-valued or spin representations; to the representations of the symmetric group and the analysis of their direct products; to the crystallographic groups; and to the Lorentz group and the concept of semi-vectors (as developed by Einstein and Mayer). —*Extract from Preface.*

380 pages, 8vo, cloth, $5.00

THE JOHNS HOPKINS PRESS · BALTIMORE 18

NUMERICAL MATHEMATICAL ANALYSIS

By JAMES B. SCARBOROUGH

"A valuable feature of the book is the excellent collection of examples at the end of each chapter. . . . The book has many admirable features. The explanations and derivations of formulae are given in detail. . . . The author has avoided introducing new and complicated notations which, although they may conduce to brevity, are a serious stumbling block to the reader. The typography and paper are excellent."
—*American Mathematical Monthly.*

430 pages, 25 figures, crown 8vo, buckram $5.50

TABLES OF $\sqrt{1-r^2}$ AND $1-r^2$ FOR USE IN PARTIAL CORRELATION AND IN TRIGONOMETRY

By JOHN RICE MINER, Sc. D.

These tables fill a want long felt by practical workers in all branches of statistics. Everyone who uses the method of correlation has wished for tables from which the probable error of a coefficient of correlation could be obtained *with accuracy*. Similar tables to this have existed on a small scale, but never before have there been available tables of $1-r^2$ and $\sqrt{1-r^2}$ to 6 places of decimals, and 4 places in the argument. Not only are these tables of great usefulness in getting the probable error of a correlation coefficient, but also they have what will perhaps be their chief value in the calculations involved in the method of partial or net correlation. It is safe to say that these tables reduce the labor involved in this widely used statistical method by at least one-half.

50 pages, 8vo, cloth, $1.50

THE JOHNS HOPKINS PRESS · BALTIMORE 18